Author:
Titterton, Alexander S

Title:
Novel Searches for NMSSM Signatures with Low Missing Transverse Energy with the CMS Detector at the LHC

University of BRISTOL

UNIVERSITY OF Southampton

Novel Searches for NMSSM Signatures with Low Missing Transverse Energy with the CMS Detector at the LHC

Alexander Titterton

A dissertation submitted to the University of Bristol and the University of Southampton in accordance with the requirements of the degree of PhD in the Faculty of Science, School of Physics, Bristol and Faculty of Physical Sciences and Engineering, School of Physics and Astronomy, Southampton.

August 2019

$\sim$ 38,000 words

University of Bristol, University of Southampton

Abstract

This thesis investigates scenarios which may arise in the Next-to-Minimal Supersymmetric Standard Model (NMSSM) whereby the Lightest Supersymmetric Particle (LSP) may in fact have a mass of only a few GeV whilst still evading current minimal Supersymmetry (SUSY) search efforts. In these NMSSM models a Next-to-Lightest Supersymmetric Particle (NLSP) decays to an SM-like Higgs boson and an LSP, with sufficiently small mass gaps such that any missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) is highly suppressed.

Firstly these scenarios are interpreted in the context of an existing $E_{\mathrm{T}}^{\mathrm{miss}}$-based analysis, considering an all-hadronic final state, in order to gauge the current sensitivity which may be attained using general purpose SUSY search techniques.

Experimental sensitivity is then extended by considering and modifying a novel analysis, which focuses on the identification of highly boosted Higgs bosons decaying to bottom quark pairs. This analysis utilises the full datasets from proton-proton collisions at a centre-of-mass energy of 13 TeV for run years 2016 and 2017 at the Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC), corresponding to respective integrated luminosities of 35.9 fb^{-1} and 41.5 fb^{-1}.

Acknowledgements

Having spent the last four years based at the University of Southampton, the Rutherford Laboratory, CERN and the University of Bristol, an enormous number of people I have had the pleasure of meeting, socialising and working with along the way deserve my gratitude. Without the encouragement and support of many of these people, this thesis would not have been possible.

Firstly, to my supervisors Dr. Henning Flaecher, Prof. Stefano Moretti and Prof. Claire Shepherd-Themistocleous I give my most profound thanks. The opportunity to undertake this PhD under your supervision as well as the advice, support and patience I have received throughout deserves my utmost gratitude.

Furthermore I extend my thanks to my colleagues at the University of Bristol, University of Southampton and Rutherford Appleton Laboratory, for both academic support and social company. To Martisse, for the most fantastic support throughout many aspects of my academic work, you deserve a great deal of thanks. To Joe, my thanks for the patience and assistance with the experimental analysis as well as plenty of laughs along the way. To Simon and Azaria, you made my time spent in Southampton both hilarious and rewarding, for which I am eternally grateful. To Mark, my thanks for making the time we spent sharing a flat at CERN such a great one.

To my fellow PhD students at these institutions and others around the world whose acquaintances I have made through CERN and many international conferences, I would like to thank you all for allowing me to share this adventure with you, for making the last four years the most fantastic experience.

To my physics teacher Mr Darke, your wit, support and wisdom played a great role in catapulting me onto the trajectory which led to me to this point in my career, I cannot thank you enough.

A very special mention is reserved for my favourite Finn, Jaana Heikkilä. Our physics discussions, times spent at CMS, our road-trips, failed band practice sessions, booze-ups and visits to Ikea are memories I will forever cherish, and our friendship means the world to me.

My close friends outside of academia deserve my utmost gratitude. To Daniel and Steven, our friendship over the many years we've known each other is something for which I feel very lucky. You have both helped keep me grounded during the ups and downs of the PhD with our Italian road trip and many other holidays. Steven, our coffee shop physics chatter and our ambitious-but-rubbish garage science experiments twelve years ago helped set me on the path towards this PhD, for which I am very thankful.

Finally, the greatest thanks of all must go to my family. To my wonderful partner Sophie; the love, support and almost endless patience you have given me during times of joy and times of stress has meant everything to me, I cannot thank you enough. To my parents, I would need another 200 pages in order to even attempt to properly convey my gratitude. Over the last 26 years you have provided me with unwavering and unconditional support, love and encouragement, giving me a childhood not one detail of which I would change for the world.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations and Code of Practice for Research Degree Programs of both the University of Bristol and the University of Southampton. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree. Accordingly any work done with the assistance of or in collaboration with others is indicated as such in the text. Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol or the University of Southampton. The dissertation has only been presented to the University of Bristol and the University of Southampton for attainment of a joint degree and has not been presented to any other university for examination either in the United Kingdom or overseas.

Signed

Date

Chapter 2 is formed primarily of an overview of the theoretical particle physics material upon which the work in this thesis is based. Whilst none of this theoretical background material is the author's own work, the mass scans presented in Section 6.1 were developed by the author in collaboration with Ulrich Ellwanger. These mass scans and the work shown in Section 6.2 were published in Ref [1], which was written by the author.

Chapter 3 and Chapter 4 respectively contain an overview of the Large Hadron Collider and Compact Muon Solenoid experiment and of the software tools utilised for the analysis work presented in this thesis. Chapter 5 details the specific observables considered in the phenomenological and experimental works in this thesis, including common objects and any analysis-specific differences in definitions thereof. These chapters do not include any original work by the author.

The author's contribution to the global CMS experimental effort involves primarily work on the Level-1 Trigger, including positions of responsibility during data-taking periods, as well as offline software development. The author has completed many shifts in the CMS control room and has spent numerous weeks as an on-call detector expert for both the Level-1 Trigger and the Layer-2 Calorimeter trigger.

Chapter 6 involves a phenomenological reinterpretation of an existing CMS analysis, and is the author's own work. Along with the work shown in Sections 6.1 and 6.2, Chapter 6 is taken mostly from Ref [1], which is the author's own work.

The work presented in chapter 7 builds on the analysis documented in Ref. [2] which was developed by CMS colleagues at Bristol and RAL. The work presented in this thesis adopts the same general analysis strategy but introduces modifications to the event selection to better target the theoretical models under investigation. This, as well as comprehensive work on background estimation and uncertainties, was performed by the author.

The software framework used for the analysis was originally developed by CMS colleagues from Bristol and RAL. All final results and figures presented in this

chapter were obtained by the author, with any unmodified work identified as such in the text.

Contents

List of Figures

List of Tables

1 | Introduction

The Standard Model (SM) [3–8] of particle physics has been incredibly successful in modelling interactions between fundamental particles at centre-of-mass energies of up to 8 TeV. No significant excesses have been observed during of Run I of the Large Hadron Collider at CERN, spanning from 2008-2012, nor in Run II which took place in 2015-2018. However, as is detailed further in Chapter 2, the Standard Model is not infallible.

The 2012 discovery of a particle decaying via two channels – to a photon pair ($\gamma\gamma$) and also to two Z bosons which then decayed to give a four-lepton final state – with properties consistent with those of an SM-like Higgs Boson [6, 7] of mass 125 GeV, does indeed provide further evidence of the strong success thus far of the SM. However the SM alone is not able to naturally motivate a Higgs Boson mass on the electroweak scale of such order. This is since contributions from fermion loops, dominated by the top quark due to its high mass, give rise to quadratically divergent contributions to the Higgs mass.

Additionally, understanding astrophysical measurements such as the Dark Matter relic density [9], as well as phenomena such as baryogenesis [10] and the lack of right-handed neutrinos [11] requires at the very least extensions to the Standard Model, if not a completely different theory. No particles in the SM have properties consistent with that of a DM candidate, with the SM unable to motivate why there is an abundance of matter in the universe compared with antimatter.

The benefits of such a theory as Supersymmetry (SUSY), which will be discussed further in Chapter 2, are many-fold. Firstly SUSY offers a more natural solution to the Hierarchy Problem [12], since the Higgs boson mass of 125 GeV is not well motivated in the SM. Furthermore a light, stable supersymmetric particle may provide a suitable Dark Matter (DM) candidate [13]. In the search for a Grand Unified Theory (GUT) it is found in the case of the Minimal Supersymmetric Standard Model (MSSM) the coupling constants for strong and electroweak inter-

actions meet at the grand unified energy (GUT scale), around 10^{16} GeV. Crucially however, this is not the case in the SM [14].

Situated on the Franco-Swiss border near Geneva, at the foot of the Jura mountains, the Large Hadron Collider (LHC) provides excellent apparatus with which to explore theoretical models which rival and extend upon the Standard Model. Having achieved one of its primary goals – the discovery of a Higgs Boson – the main objective of the LHC is now to study particles resultant from high energy proton collisions, to search for signatures of possible new physics.

Whilst a result which disagrees significantly with the SM expectation would certainly warrant much ado, a challenge is finding a theory which agrees with the observed yields across all measurements performed on the data to a better extent than the Standard Model prediction. One such theory which extends the Standard Model and addresses its shortcomings is SUSY. However a consequence of this, as will be discussed further in chapter 2, is the introduction of a bosonic degree of freedom for each fermion, and vice-versa.

This thesis begins by introducing and motivating the Next-to-Minimal Supersymmetric Standard Model (NMSSM) as an extension to the SM, discussed in chapter 2. The LHC, with particular emphasis on the CMS detector and its Trigger system, are detailed in chapter 3, followed by a comprehensive description of the various software used in order to generate simulated Monte Carlo events, from the initial matrix element-level parton scattering calculations through to the simulation of the detector response and object reconstruction, in chapter 4.

We continue by motivating and exploring particular NMSSM scenarios whereby heavier supersymmetric particles decay producing, per event, two Standard Model-like Higgs Bosons and two stable, neutral supersymmetric particles which are themselves invisible to the detector. The work presented in this thesis focusses specifically on the case where these invisible, stable particles are comparably light and have small transverse momenta, meaning the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) – usually a feature by which many SUSY models may be distinguished from SM

background processes – is highly suppressed, therefore allowing this model to potentially avoid existing search efforts at a hadron collider. This differs from the typical MSSM scenario, in which a relatively heavy LSP is produced with larger transverse momentum, leading to higher $E_{\mathrm{T}}^{\mathrm{miss}}$.

With an aim to assess the sensitivity of existing SUSY searches to these low-$E_{\mathrm{T}}^{\mathrm{miss}}$ scenarios a general purpose jets+$E_{\mathrm{T}}^{\mathrm{miss}}$ search [15] was recast in a phenomenological context. In order to improve sensitivity to the above mentioned NMSSM scenarios we further develop the analysis in Ref [2] to gain access to previously uncovered regions of parameter space, by introducing extra event yield binning in the $H_{\mathrm{T}}^{\mathrm{miss}}$ variable, a hadronic jet-based quantity similar to $E_{\mathrm{T}}^{\mathrm{miss}}$. This analysis, designed around searching for similar SUSY processes but with a light scalar Higgs Boson, is interpreted and advanced in the context of the NMSSM scenarios under investigation in this thesis. The analysis centres around a double-b-tag boosted decision tree (BDT) discriminator, used to target boosted Higgs Bosons decaying to a bottom quark-antiquark pair, where the Higgs Boson momentum is sufficiently high that resolving *both* of the bottom quarks into individual hadronic jet cones may not be efficient or even possible.

In order to tailor such a search toward the particular scenarios considered in this thesis, modifications are made to take advantage of the extra $E_{\mathrm{T}}^{\mathrm{miss}}$ generated when increasing the LSP mass, since in the case of a high-momentum NLSP decaying into a SM-like Higgs boson and an LSP, a heavier LSP will inherit a larger fraction of the NLSP momentum. This allows for greater sensitivity in areas of parameter space which are not adequately accessible by either the double-b-tag analysis in its original form or the phenomenology work in Chapter 6.

2 | Theory

2.1 The Standard Model

The SM [3–8] has achieved enormous success in describing the interactions between fundamental particles via Quantum Field Theory (QFT), in particular perturbation theory. SM forecasts based upon Quantum Electro-Dynamics (QED) in particular have attained agreement with observations in data to within incredible precision, for example predicting the value of the electromagnetic fine structure constant, α_S, to one part in 10^8 [16].

As a model the SM has remained unchanged since its completion in the 1970s, and in this time has been further bolstered by the discoveries of the charm quark [17], tau lepton [18], bottom quark [19], W boson [20], Z boson [21] and the top quark [22, 23]. Finally in 2012, discovery of a particle consistent with a SM Higgs boson was confirmed independently by the CMS [24] and ATLAS [25] collaborations at CERN, completing the manifest of particles implied by the SM.

2.1.1 Standard Model Gauge Theory

The Standard Model is an $SU(3)_c \times SU(2)_L \times U(1)_Y$ gauge theory. The $SU(3)_c$ gauge group represents Quantum Chromo-Dynamics (QCD), which describes the strong nuclear force, is a vector-like theory, coupling to both left- and right-handed states. The $SU(2)_L \times U(1)_Y$ gauge group on the other hand represents electroweak theory, which describes electromagnetic and weak force interactions, and is a chiral theory [16].

The SM contains fermions:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L \qquad \text{(left-handed quarks)};$$

$$u_R, d_R, c_R, s_R, t_R, b_R \qquad \text{(right-handed quarks)};$$

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}_L \quad \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L \quad \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L \qquad \text{(left-handed leptons)};$$

$$e_R, \mu_R, \tau_R \qquad \text{(right-handed leptons)};$$

gauge bosons:

$$g, \gamma, Z, W^{\pm};$$

and a scalar Higgs boson:

$$H. \tag{2.1}$$

The left-handed fermion multiplets transform as $SU(2)$ doublets, whilst the right-handed fermions transform as $SU(2)$ singlets. The left- and right-handed matter fields, as well as the gauge bosons and the Higgs field may be written as follows:

Quarks:

$$\begin{aligned} Q = \begin{pmatrix} U_L \\ D_L \end{pmatrix} : &\quad (3, 2, -\tfrac{1}{6}), \\ U_R^\dagger : &\quad (3, 1, -\tfrac{2}{3}), \\ D_R^\dagger : &\quad (3, 1, \tfrac{1}{3}), \end{aligned}$$

Leptons:

$$L = \begin{pmatrix} \nu_L \\ E_L \end{pmatrix} : \quad (1, 2, -\tfrac{1}{2}),$$
$$E_R^\dagger : \quad (1, 1, 1),$$

Gauge Bosons:

$$g : \quad (8, 2, 0),$$
$$\gamma : \quad (1, 1, 0),$$
$$W^\pm : \quad (1, 3, 0),$$
$$Z : \quad (1, 1, 0),$$

Higgs:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} : \quad (1, 2, \tfrac{1}{2}), \tag{2.2}$$

with the $SU(3) \times SU(2) \times U(1)$ quantum numbers written in parentheses for each respective field. Here we note that Q, U, D, L, ν and E have flavour indices suppressed and are thus treated as vectors in family space.

An important phenomenon in the SM is Electro-Weak Symmetry Breaking (EWSB). EWSB is the spontaneous symmetry breaking $SU(2)_L \times U(1)_Y \to U(1)_Q$, allowed by the inclusion in the SM of the Higgs field, with the electromagnetic and weak forces being unified above a unification energy of around 100 GeV. Following EWSB, the Higgs field acquires a vacuum expectation value, v, with the W and Z bosons gaining mass terms. Fermions also gain mass terms via Yukawa couplings Y_f between themselves and the Higgs field [16].

Figure 2.1: One-loop corrections to the Higgs potential V, where the fermion loop term will have opposite sign with respect to the scalar loop term.

2.2 Limitations of the Standard Model

The Standard Model does a remarkable job in describing the fundamental behaviour of particles at energy scales currently investigated, however it does come with caveats. One of the main problems with the Standard Model is known as the Hierarchy Problem [13], and may be described as follows.

Taking for example a theory involving some scalar particle with self-coupling $|\lambda_H|^2$, which couples to fermions via Yukawa couplings with strength λ_f. Referring to this scalar particle as a Higgs boson, corrections to the Higgs potential may be found at one-loop order, which arise from the diagrams shown in Figure 2.1.

These diagrams give the following divergent contribution to the Higgs potential, leading to fine tuning of the order 10^{16} required in order to obtain the observed Higgs mass of 125 GeV:

$$\Delta M_{\text{Higgs}} \sim (|\lambda_H|^2 - |\lambda_f|^2) H^2 \Lambda^2, \tag{2.3}$$

where $|\lambda_f|^2$ represents a sum over the loop term for each fermion flavour.

This quadratic mass contribution becomes problematic when considering the large coupling to the top quark, due to it being much heavier than the other fermions. This leads to a large divergence, for which we require the $|\lambda_H|^2$ and $|\lambda_f|^2$ terms to cancel.

2.3 Supersymmetry

Introducing supersymmetry has the advantage of sending $|\lambda_H|^2, |\lambda_f|^2 \to 0$, cancelling these divergent contributions to the Higgs mass [12]. However, as a consequence, we must introduce a fermionic degree of freedom for each boson and a bosonic degree of freedom for each fermion, giving rise to supersymmetric partners, *superpartners*, of Standard Model particles. The magnitude of the coupling terms of the SM fermions equal those of their respective superpartners, but with the difference that since the superpartner fields are bosonic the coupling terms in the loop diagrams contributing to the Higgs mass have opposite sign compared with the SM fermions. This therefore allows for the quadratic terms contributing to ΔM_H to cancel.

We consider an anti-commuting spinor operator Q, which generates SUSY transformations of the form

$$\begin{aligned} Q|\text{Boson}\rangle &= |\text{Fermion}\rangle, \\ Q|\text{Fermion}\rangle &= |\text{Boson}\rangle. \end{aligned} \tag{2.4}$$

Taking SUSY to be a theory which, like the SM, allows parity-violating interactions, and so features chiral fermions, we must have generators Q and $Q^\dagger$ which obey commutation and anticommutation relations as follows:

$$\{Q, Q^\dagger\} \sim P^\mu, \tag{2.5}$$

$$\{Q, Q\} = \{Q^\dagger, Q^\dagger\} = 0, \tag{2.6}$$

$$[P^\mu, Q] = [P^\mu, Q^\dagger] = 0, \tag{2.7}$$

where the spinor indices on the Q terms are suppressed. Here P^μ represents the four-momentum operators, which generate space-time translations, with the eigenvalues being conserved four-momenta [26].

2.3.1 Superpartners to SM Particles

All particles in a supersymmetric theory exist in irreducible algebra representations known as supermultiplets, with each supermultiplet containing a fermionic and a bosonic state; superpartners of each other. Considering a supersymmetric extension of the SM, examples of which will be explored further in this chapter, each of the SM fundamental particles belongs to either a gauge or chiral supermultiplet, depending on its properties.

Left- and right-handed SM fermions transform differently, and so must be part of a chiral supermultiplet [13]. Since the superpartner of a particle must have spin differing by $\frac{1}{2}$ from that of said particle, the superpartner to each of the quarks and leptons must have either spin-0 or spin-1. However, in the case of the left-handed leptons, being $SU(2)$ doublets as shown in Equation (2.2), the superpartners cannot have spin-1, as these would be gauge bosons which transform according to the three-dimensional adjoint representation of $SU(2)_{\mathrm{L}}$, not the doublet representation [26].

Additionally, the quarks form triplets of the $SU(3)_{\mathrm{c}}$ colour gauge group, whereas gauge bosons exist only in the octet representation, exactly as in the SM, with e.g. the gluon consisting of eight independent colour states [27, 28]. As such we find that, similar to the leptons, the superpartners of the quarks must also be spin-0 particles. These spin-0 partners to the leptons and quarks acquire a prefix "s": for example, the superpartner to the top quark is the stop, and the superpartner to the electron is known as the selectron.

The remaining chiral supermultiplets constitute the Higgs sector. However, in order to avoid gauge anomalies [13, 26], we must in fact have *two* Higgs chiral supermultiplets, H_u and H_d, known as Higgsinos, in order to give masses to the up- and down-type fermions. Summarising, we present the (s)lepton, (s)quark and Higgsino components of a supersymmetric extension to the SM in Equation (2.8).

Quarks:

$$\begin{aligned}\begin{pmatrix} U_L \\ D_L \end{pmatrix}, \begin{pmatrix} \tilde{U}_L \\ \tilde{D}_L \end{pmatrix} &: \quad (3, 2, -\tfrac{1}{6}), \\ U_R^\dagger, \tilde{U}_R^* &: \quad (3, 1, -\tfrac{2}{3}), \\ D_R^\dagger, \tilde{D}_R^* &: \quad (3, 1, \tfrac{1}{3}),\end{aligned}$$

Leptons:

$$\begin{aligned}\begin{pmatrix} \nu_L \\ E_L \end{pmatrix}, \begin{pmatrix} \tilde{\nu}_L \\ \tilde{E}_L \end{pmatrix} &: \quad (1, 2, -\tfrac{1}{2}), \\ E_R^\dagger, \tilde{E}_R^* &: \quad (1, 1, 1),\end{aligned}$$

Higgs:

$$\begin{aligned}H_u = \begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix}, \begin{pmatrix} \tilde{H}_u^+ \\ \tilde{H}_u^0 \end{pmatrix} &: \quad (1, 2, \tfrac{1}{2}), \\ H_d = \begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix}, \begin{pmatrix} \tilde{H}_d^0 \\ \tilde{H}_d^- \end{pmatrix} &: \quad (1, 2, -\tfrac{1}{2}),\end{aligned} \tag{2.8}$$

where as before U, D, E and ν are treated as vectors in family space, with indices suppressed.

The gauge bosons, being spin-1 vector bosons, naturally reside along with their respective spin-$\frac{1}{2}$ superpartners in gauge supermultiplets. These superpartners acquire the suffix "-ino", for example the superpartner to the gluon being the gluino.

Furthermore the massive SM gauge boson fields are changed: rather than $W^\pm$, Z and γ, we now have W^+, W^0, W^- and B^0, with respective superpartners $\tilde{W}^+$, $\tilde{W}^0$, $\tilde{W}^-$ and $\tilde{B}^0$: three *winos* and one *bino*. Following EWSB, the W^0 and B^0 fields mix to give mass eigenstates Z^0 and γ, with corresponding gaugino mixtures

of $\tilde{W}^0$ and $\tilde{B}^0$ giving a *zino* and a *photino*. Equation (2.9) below summarises the gaugino sector of a supersymmetric extension to the SM [13, 26].

Gauge Bosons:

$$\begin{aligned} g, \tilde{g} : &\quad (8, 1, 0), \\ W^{\pm}, W^0, \tilde{W}^{\pm}, \tilde{W}^0 : &\quad (1, 3, 0), \\ B^0, \tilde{B}^0 : &\quad (1, 1, 0), \end{aligned} \tag{2.9}$$

2.3.2 Broken Supersymmetry

Crucially, since the squared mass operator $-P^2$ commutes with Q and $Q^\dagger$, the particles inhabiting the same supermultiplet must, in the unbroken SUSY theory, have the same mass eigenstates. Given that *none* of the superpartners have as yet been found, let alone an abundance of selectrons with mass 0.511 MeV– in the presence of which the decay width of the Z boson would be double – it is theorised that below some energy scale SUSY is a broken symmetry, allowing for the superpartners to have masses much higher than their Standard Model counterparts [13].

In order for the relationships between dimensionless couplings to hold between the broken and unbroken SUSY theories, critical for the canceling needed to solve the hierarchy problem, we consider "soft" SUSY breaking. In this case an effective Lagrangian can be written as

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{Soft}}, \tag{2.10}$$

where the former term is SUSY-invariant and contains the gauge and Yukawa interaction terms, and the latter contains only mass terms and couplings with pos-

itive mass dimension. This condition ensures that dimensionless SUSY-breaking couplings are absent, which allows for the cancellation of quadratically divergent terms contributing to the masses of scalar particles [13, 26].

2.4 The Minimal Supersymmetric Standard Model

A supersymmetric theory which deviates too far from the Standard Model at low energies would not necessarily be favourable or sensible, given the Standard Model's success thus far at describing subatomic physics at energy scales up to that which is currently accessible by collider experiments. Therefore we introduce the Minimal Supersymmetric Standard Model, MSSM, which contains enough supersymmetry components to satisfy Equation 2.3 whilst at low energy appearing similar to the Standard Model.

2.4.1 Particle Contents of the MSSM

The MSSM contains the superfields shown in Equations (2.8) and (2.9). In addition to the SM quarks, leptons and gauge bosons and Higgs boson, the MSSM includes an assortment of SUSY particles.

Firstly, the MSSM contains four neutralinos, denoted χ_1^0, χ_2^0, χ_3^0, χ_4^0; neutral Majorana fermions formed via the mixing of the neutral Higgsino and neutral gaugino fields [13].

The charged Higgsino and wino components also mix, creating charginos. These electrically charged fermions are usually denoted $\chi_1^\pm$ and $\chi_2^\pm$. Additionally sleptons, the scalar superpartners of the leptons, also appear. Neither charginos nor sleptons are of particular direct relevance to the specific decay channels or final states considered in this thesis, and as such their phenomenology is not discussed further.

Finally and importantly, squarks and gluinos, the superpartners of the quarks and gluons respectively, are also found in the MSSM. In many models, such as those considered in this thesis, the first two generations of squarks are degenerate in

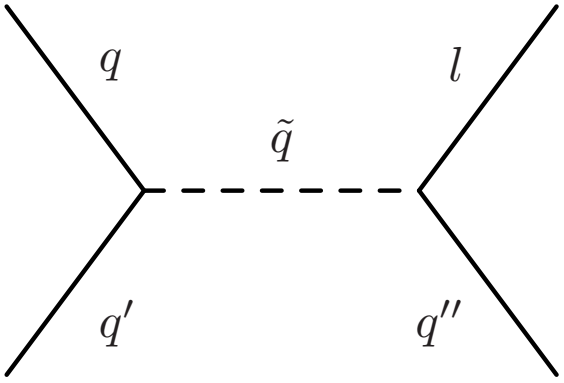


Figure 2.2: Example interaction by which interactions between SM particles may produce a single squark, which then decays into SM particles.

mass, complying with constraints from flavour changing neutral currents (FCNC), with the stop and sbottom squarks allowed independent masses. Notably the gluino, being a colour-octet Majorana fermion, cannot mix with any other particle in the MSSM. As such, its mass term in the Lagrangian is unaffected by any other particle mass or coupling term.

2.4.2 *R*-Parity

The full MSSM Lagrangian contains terms which, whilst remaining gauge invariant, violate lepton number or baryon number [13]. A serious consequence of this is illustrated in Figure 2.2. In this case it may be possible for single supersymmetric particles to be produced from SM particles, and also to decay entirely into SM particles.

Figure 2.3 shows an example of how a squark decaying into SM particles would allow for squark-mediated proton decay. This allows for decays such as $p \rightarrow e^+ + \pi^0$, and implies that the proton decay lifetime be very short, of the order of 10^{-2} seconds [13].

However, experimental measurements [29] have placed lower bounds on the decay lifetime of the proton $\tau > 10^{33}$ years, suggesting that any proton decay interactions, if allowed at all, must be very highly suppressed. Thus, in order to find a natural way to resolve this issue, a new symmetry is introduced: R-parity.

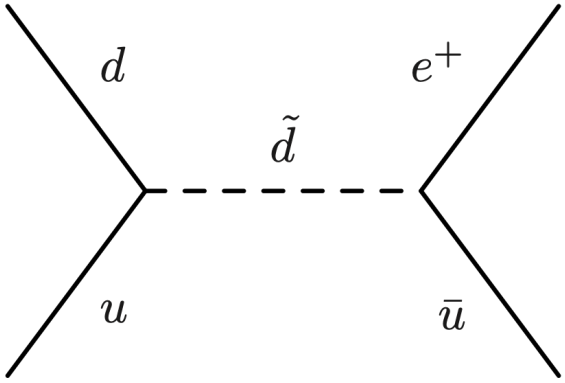


Figure 2.3: Example interaction by which proton decay may be mediated by a squark, giving a short proton lifetime.

We define R, a multiplicative quantum number, as follows [13]:

$$R := (-1)^{3B+L+2s}, \tag{2.11}$$

where B, L and s represent baryon number, lepton number and spin respectively, noting that $R = -1$ for SUSY particles and $R = +1$ for SM particles. A possible origin of R-parity is from a continuous $B - L$ gauge symmetry, spontaneously broken at energies higher than is currently accessible by collider experiments [30]. Imposing conservation on baryon or lepton numbers also solves the issue of proton decay, however this then leads to R-parity-violating scenarios; the MSSM therefore requires the introduction of additional symmetry to avoid such issues. For the remainder of this thesis we consider only the case where R-parity is upheld.

A result of the enforcement of this extra symmetry is that the lightest supersymmetric particle (LSP) must be stable, since it cannot decay into lighter SM particles without violating R-parity. The LSP is usually formed via weak interactions in SUSY decay cascades and gives rise to *missing energy*, since it carries momentum but is invisible to direct detection within a particle detector. In the context of the MSSM, the role of LSP is often played by the lightest of the neutralinos, χ_1^0.

Furthermore, a stable neutralino LSP is also highly motivated within Dark Matter (DM) models, forming a candidate for cold DM. Observations of galactic rotation imply the existence of DM, in the absence of a problematic and elusive modified

gravity-based solution [9]. SM neutrinos alone cannot account for the observed DM relic density, and gravitinos form hot DM which is inconsistent with other experimental measurements [31]. Neutralinos however remain a viable DM candidate.

2.4.3 The 'μ Problem'

The MSSM however is not without its own drawbacks. The most notorious caveat to the MSSM is known as the *μ-problem* [12, 32]. Examining the MSSM Lagrangian we find the following superpotential:

$$W_{\text{MSSM}} = \bar{u}\mathbf{y_u}QH_u - \bar{d}\mathbf{y_d}QH_d - \bar{e}\mathbf{y_e}QH_d + \mu\hat{H}_u \cdot \hat{H}_d + ... \tag{2.12}$$

where $\mathbf{y_u}$, $\mathbf{y_d}$ and $\mathbf{y_e}$ are the 3×3 Yukawa coupling matrices in family space,

$$H_u := \begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix} \qquad \text{and} \qquad H_d := \begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix}, \tag{2.13}$$

with H_u and H_d being components of chiral superfields which additionally contain fermionic $SU(2)$ doublets ψ_u and ψ_d respectively. As mentioned in Section 2.4.1 charged components of these doublets, along with the fermionic superpartners of the W^+ and W^- bosons, form the chargino sector of the MSSM.

These fields may have supersymmetric mass terms in the MSSM Lagrangian: a positive mass squared μ^2 term for $|H_u|^2$ and $|H_d|^2$, along with a Dirac mass μ for the $SU(2)$ doublets ψ_u and ψ_d. Additionally a soft SUSY breaking mass term $B\mu H_u H_d$ may appear, where B also has mass dimension.

This μ parameter has dimensions of mass, and the only "natural" values are $\mu = 0$ and $\mu \sim M_{\text{Planck}}$. However for the following phenomenological reasons μ cannot

simply vanish, nor may it take very large values. Firstly, given the fermionic $SU(2)$ doublets ψ_u and ψ_d contain charged components, a Dirac mass term μ is required. However, as discussed above these make up part of the chargino sector of the MSSM, with LEP searches placing a lower bound of around 103 GeV on the chargino mass [33]. Studies of the chargino mass matrix [34] have shown that regardless of the values taken by other parameters, we must have $|\mu| \gtrapprox 100$ GeV in order for this LEP constraint to be satisfied.

Moreover, the trivial solution $\mu = 0$ cannot be allowed, since this unavoidably leads to a forbidden massless axion [32], stemming from a Peccei-Quinn (PQ) symmetry.

Additionally, were the neutral components of H_u and H_d to vanish at the minimum, EWSB, necessary to give masses to the quarks and leptons, would not occur. Therefore in order to generate non-zero masses for leptons as well as up- and down-type quarks via the Higgs mechanism, these components must not vanish. Furthermore we should have μ of the order of the Z boson mass, in order to provide a vacuum expectation value (vev) v for the Higgs potential in agreement with the W and Z boson masses:

$$v^2 := v_u^2 + v_d^2 = \frac{4m_W^2}{g^2} \simeq (246 \text{ GeV})^2, \tag{2.14}$$

where m_W is the W boson mass and g the weak isospin coupling. Equation (2.14) in combination with:

$$m_Z^2 = \frac{2(m_{H_d}^2 - m_{H_u}^2 \tan^2\beta)}{\tan^2\beta - 1} - 2|\mu|^2, \tag{2.15}$$

$$\sin 2\beta = \frac{2B\mu}{2\mu^2 + m_{H_u}^2 + m_{H_d}^2}, \tag{2.16}$$

lead to the condition that μ^2 must not be much larger than m_Z^2 without requiring large amounts of fine-tuning [12, 32, 34].

Finally, we consider the solution in which μ takes very large values, of the order of the Planck mass M_{Planck}. In order to have EWSB, necessary in order to generate

lepton and quark masses, the Higgs potential must be unstable at the origin; where $H_u = H_d = 0$. Soft SUSY breaking (SSB) mass terms for these fields may allow for an unstable origin, so long as they are not dominated by the μ-induced positive mass-squared values for H_u and H_d, implying an upper limit of $|\mu| \leq M_{\text{SUSY}}$ [13, 34], where M_{SUSY} is the mass scale at which soft-SUSY is broken.

To summarise, the MSSM features a μ parameter, which has dimensions of mass and yet must be set by hand to be at the electroweak scale, around 100GeV–1TeV. However, the only “natural" values μ may take are either zero or around the Planck mass: both being disallowed and far from the electroweak scale. Whilst this is not a fatal problem for the MSSM it raises questions as to why the μ scale should happen to fall so close to the electroweak scale. Instead however, we consider an extended version of the supersymmetric theory, the Next-to-Minimal Supersymmetric Standard Model, which does not feature a dimensionful μ term and so has no “”μ problem".

2.5 The Next-to-Minimal Supersymmetric Standard Model

The Next-to-Minimal Supersymmetric Standard Model, NMSSM, presents a natural solution to the μ-problem, since it does not contain dimensionful parameters whose values must be set by hand. This is achieved by introducing a gauge singlet field $\hat{S}$ which facilitates the removal of the problematic μ term [34].

Other approaches to the μ-problem include the Giudice-Masiero mechanism [35], whereby the μ term does not appear in the MSSM Lagrangian since it violates some global symmetry, as well as proposed solutions which require R-parity to be violated [36]. However, these will not be discussed further in this thesis.

2.5.1 $\mathbb{Z}_3$-invariant NMSSM

Following from the MSSM superpotential, given in Equation (2.12), we see below in Equation (2.17) the same terms in the NMSSM superpotential, where the λ term has replaced the mischievous μ-term [34].

$$W_{\text{NMSSM}} = \text{Yukawa Couplings } (q, l^+, l^- \text{ masses}) + \lambda S H_u H_d + \frac{1}{3}\kappa \hat{S}^3 + ... \tag{2.17}$$

The λ term in Equation (2.17) is a dimensionless constant and S gives rise to a singlino; the SUSY counterpart of a singlet Higgs boson. Thus the Higgs sector of the NMSSM extends that of the MSSM, resulting in the inclusion of seven Higgs bosons, including superpartners, as opposed to five, since S is a chiral superfield, with scalar component $\hat{S}$ and spin-$\frac{1}{2}$ component $\tilde{S}$. Furthermore, in contrast to the MSSM, the lightest Higgs boson in the NMSSM is not necessarily an SM-like Higgs boson, and can be very light [34, 37]. The respective mass eigenstates present in the MSSM and NMSSM are compared in Table 2.1.

Here we consider the "$\mathbb{Z}_3$-invariant" NMSSM superpotential which possesses only cubic terms, so-called since we encounter an accidental $\mathbb{Z}_3$ symmetry, since the Lagrangian remains unchanged when multiplying each field by $e^{2\pi i/3}$. Since the λ and κ factors multiplying these cubic terms are dimensionless, we obtain a more natural solution.

The full "general" NMSSM contains in addition terms linear and quadratic in S. However, these once again lead to dimensionful parameters whose values must be set by hand [34, 38], thus eroding the naturalness of the NMSSM as a solution to the μ-problem. Therefore, for the remainder of this thesis we will focus solely on the "$\mathbb{Z}_3$-invariant" NMSSM, denoting it simply as the NMSSM in keeping with the literature [34, 37, 38].

Particle	MSSM	NMSSM
Higgs Bosons	h^0, H^0, A^0, $H^\pm$	H_1, H_3, H_3, A_1, A_2, $H^\pm$
Squarks	$\tilde{u}_L, \tilde{u}_R, \tilde{d}_L, \tilde{d}_R$	(same)
	$\tilde{c}_L, \tilde{c}_R, \tilde{s}_L, \tilde{s}_R$	(same)
	$\tilde{t}_1, \tilde{t}_2, \tilde{b}_1, \tilde{b}_2$	(same)
Sleptons	$\tilde{e}_L, \tilde{e}_R, \tilde{\nu}_e$	(same)
	$\tilde{\mu}_L, \tilde{\mu}_R, \tilde{\nu}_\mu$	(same)
	$\tilde{\tau}_L, \tilde{\tau}_R, \tilde{\nu}_\tau$	(same)
Neutralinos	$\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0, \tilde{\chi}_4^0$	$\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0, \tilde{\chi}_4^0, \tilde{\chi}_5^0$
Charginos	$\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$	(same)
Gluino	$\tilde{g}$	(same)
Gravitino	$\tilde{G}$	(same)

Table 2.1: Mass eigenstates of the MSSM compared with the NMSSM [39, 40].

We no longer have any dimensionful parameters which we must set by hand, meaning we have no μ-problem. It may be desirable, however, to obtain an equivalent μ term without the naturalness problems. In order to do this we may generate an effective μ term via symmetry breaking when $\hat{S}$ has a non-zero vacuum expectation value of scale μ:

$$\begin{aligned} \lambda \hat{S} \hat{H}_u \hat{H}_d \to \lambda \langle S \rangle H_u H_d = \mu_{\text{eff}} H_u H_d \\ \rightsquigarrow \mu_{\text{eff}} = \lambda \langle S \rangle. \end{aligned} \tag{2.18}$$

The final term in Equation (2.17) represents a trilinear coupling S^3, along with a dimensionless factor κ. One drawback of the NMSSM is that the $\mathbb{Z}_3$ is broken by the vacuum expectation value of S [38], leading to domain walls in the early universe which would have been observed in the cosmic microwave background. One solution to this is to set the κ term in Equation (2.17) to zero, however this generates an additional $U(1)'$ PQ symmetry, leading to a forbidden massless axion. Numerous models have been postulated in order to remove this symmetry, notably an NMSSM variant called the new MSSM (nMSSM) [34], which in the unbroken

SUSY limit contains only the $\lambda S H_u H_d$ term. The tadpole terms, which generate one-loop Feynman diagrams with an external leg and are necessary in order to minimise the Higgs potential, are then generated when SUSY is softly broken in order to break the $\mathbb{Z}_3$ and PQ symmetries [38]. These tadpole terms are generated via the S^3 term in the NMSSM superpotential. However for the remainder of this thesis we will continue to focus only on the NMSSM, with κ taking non-zero values throughout.

2.6 Motivating NMSSM Scenarios with low $\boldsymbol{E}_{\mathbf{T}}^{\mathbf{miss}}$

After four years of proton-proton collisions at the LHC with a centre-of-mass energy of $\sqrt{s} = 7 - 8\,\text{TeV}$ and a further three with 13 TeV, searches for physics Beyond the SM (BSM) have so far not observed any significant excesses. Particularly, in searches for SUSY [41–44], this has allowed lower bounds to be placed on the masses of supersymmetric particles such as squarks, gluinos, gauginos and Higgsinos.

Recent LHC analyses such as [15], utilising the α_{T} kinematic variable [45, 46], a variable detailed in Chapter 5 designed to distinguish events featuring genuine sources of $E_{\text{T}}^{\text{miss}}$ from those with spurious $E_{\text{T}}^{\text{miss}}$, have pushed the lower bounds on the squark mass $M_{\tilde{q}}$ and gluino mass $M_{\tilde{g}}$ well in excess of 1 TeV and the mass of a neutralino LSP $M_{\tilde{X}_1^0}$ also as high as 1 TeV for certain regions of parameter space of simplified SUSY models [47–52]. These experimental limits are of course dependent upon various properties of the decay cascade such as the masses of other sparticles, the decay branching fractions and the kinematic distributions of the decay products, but have still ruled out a large area of parameter space for simplified MSSM models.

2.6.1 Theoretical Motivations

The majority of the SUSY search effort so far has relied upon the notion that an R-parity-conserving supersymmetric model is expected to generate events featuring large missing transverse energy, $E_{\text{T}}^{\text{miss}}$. In addition, long SUSY decay cascades often imply many hadronic jets with large transverse momentum, p_{T}, meaning events whose jet p_{T} scalar sum, H_{T}, is also very high.

We consider the case where the Next-to-LSP (NLSP) decays into an LSP plus a SM-like Higgs boson, H. In a scenario where the LSP carries only small momentum in this decay the $E_{\text{T}}^{\text{miss}}$ is reduced considerably. An example of such a scenario would be for a very light LSP where the NLSP mass, $M_{\tilde{\chi}_2^0}$, is just slightly above $M_{\tilde{\chi}_1^0}$

$+M_H$, since in the case of heavy squarks and gluinos decaying into an NLSP, the NLSP would have very high momentum, becoming relativistically boosted, such that the Higgs boson and LSP would be effectively collinear. The Higgs boson would then inherit most of the momentum from the NLSP compared with the much lighter LSP in this boosted regime, due to conservation of four-momentum in the NLSP rest frame.

In the case where the LSP is the singlino of the NMSSM [34], having only small couplings to sparticles, then the heavier sparticles will decay into a bino-like NLSP, which plays the rôle of the MSSM-like LSP, since in the case where the singlino LSP interacts only with other Higgs bosons in the NMSSM the squarks and gluinos cannot decay into an LSP directly. This NLSP then decays into a low momentum, "true" LSP and a Higgs boson, thus allowing for these low-$E_{\mathrm{T}}^{\mathrm{miss}}$ scenarios.

In the case where this singlino LSP interacts only via Lagrangian terms featuring a Higgs boson, without singlino mixing, such a decay route may be enforced in the NMSSM. In the MSSM however, a bino-like LSP may also be produced, with hadronic jets, directly from the decays of much heavier squark and gluinos, as well as from the decays of heavier neutralinos. In this case even a light LSP with mass of a few GeV would inherit a much larger momentum, being heavier than the light flavour quarks, thus generating large $E_{\mathrm{T}}^{\mathrm{miss}}$ once more. As such, we require the LSP to be a singlino of the NMSSM in order to motivate such low-$E_{\mathrm{T}}^{\mathrm{miss}}$ scenarios.

In the case where we have the NLSP mass $M_{\tilde{\chi}_2^0}$ not much smaller than $M_{\tilde{q}}$ or $M_{\tilde{g}}$ but still close to $M_{\tilde{\chi}_1^0}$ + 125 GeV, the now heavy LSP will gain fairly little extra momentum compared to the initially produced sparticles. This is since the smaller mass gaps in the decays will result in a lower momentum NLSP, and so the NSLP decay products will not be boosted. Additionally, these small mass gaps in the decay cascade will mean low-p_{T} jets, implying events with low H_{T} as well as low $E_{\mathrm{T}}^{\mathrm{miss}}$. This is in contrast to the typical signature of a more minimal SUSY model, whereby heavy squarks and gluinos typically decay into a lighter, invisible LSP, plus hadronic jets, leading to high H_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$.

Point	$M_{\tilde{q}}$ [GeV]	$M_{\tilde{g}}$ [GeV]	$M_{\tilde{t},\tilde{b}}$ [GeV]	Average $E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]
BP1	1000	1010	decoupled	143
BP2	1400	1410	decoupled	191
BP3	1100	900	decoupled	113
BP4	1500	1300	decoupled	153
BP5	1400	1410	$M_{\tilde{t}} = 750$	197
BP6	1100	1110	$M_{\tilde{b}} = 750$	160
BP7	1500	1300	$M_{\tilde{t}} = 750$	200
BP8	1400	1200	$M_{\tilde{b}} = 750$	158

Table 2.2: Original BPs in Ref [37].

2.6.2 Independent Mass Hierarchies

In order to explore how the proposed low-$E_{\mathrm{T}}^{\mathrm{miss}}$ NMSSM scenarios may evade current search efforts, and to develop novel search techniques in order to better probe the associated parameter space, we first consider example mass points which exhibit this low-$E_{\mathrm{T}}^{\mathrm{miss}}$ signature.

We start with eight Benchmark Points (BPs), denoted BP1, ..., BP8 and presented in Table 2.2, taken from [37]. These BPs were constructed to demonstrate mass configurations in the NMSSM which exhibit very low $E_{\mathrm{T}}^{\mathrm{miss}}$. The squark and gluino masses were chosen such that they provide a sufficiently boosted NLSP so as to give small $E_{\mathrm{T}}^{\mathrm{miss}}$ whilst not yet having been ruled out by SUSY search efforts during Run-I of the LHC, ascertained using CheckMATE [53], a software package which compares simulated events corresponding to a model with results from experimental searches. These BPs then form a basis for the discrete mass scans which will be constructed throughout this thesis.

For example, the $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions are shown in figure 2.4 for each of the eight BPs in Table 2.2, featuring a LSP mass of 3 GeV with the squark and gluino masses around 1 TeV. Also shown is a simplified MSSM-like scenario whereby squarks and gluinos decay into a stable $\tilde{\chi}_2^0$ which acts as an MSSM-like LSP, avoiding the production of Higgs bosons and thus generating higher $E_{\mathrm{T}}^{\mathrm{miss}}$. This MSSM-like scenario features the same sparticle masses as in BP1, such as 1 TeV squarks, and a 3 GeV LSP.

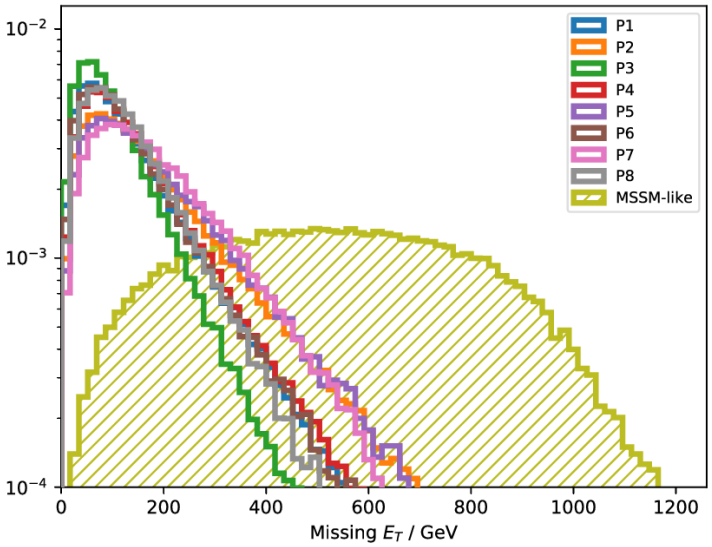


Figure 2.4: $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for the eight benchmark points defined in Ref [37] along with an MSSM-like simplified scenario.

Here, it is clear that the mean $E_{\mathrm{T}}^{\mathrm{miss}}$ is rather low for these NMSSM scenarios, more akin to that from SM processes such as fully- or semi-leptonically decaying pair-produced top quarks, as shown in Figure 2.5, or processes such as $Z \to ll$ or $W \to l\nu$, where l represents a lepton. In turn, this suggests current experimental searches concentrating on a hadronic jets plus $E_{\mathrm{T}}^{\mathrm{miss}}$ final state will likely not be optimally tuned to this type of SUSY signature.

2.6.3 Possible Decay Cascades

In order to further explore the experimental signatures of these benchmark points we examine the sets of possible decay cascades across the different mass hierarchies. In all cases, as shown in Figure 2.4, we see a suppressed $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, irrespective of the mass hierarchy between the squarks and gluino, and whether or not sbottom and stop squark masses are decoupled. Additionally this suppressed $E_{\mathrm{T}}^{\mathrm{miss}}$ spectrum requires the NMSSM, since in the MSSM the squarks and gluinos may decay directly into jets and an LSP, generating much larger values of $E_{\mathrm{T}}^{\mathrm{miss}}$ even when the LSP mass is only a few GeV, illustrated in Figure 2.4.

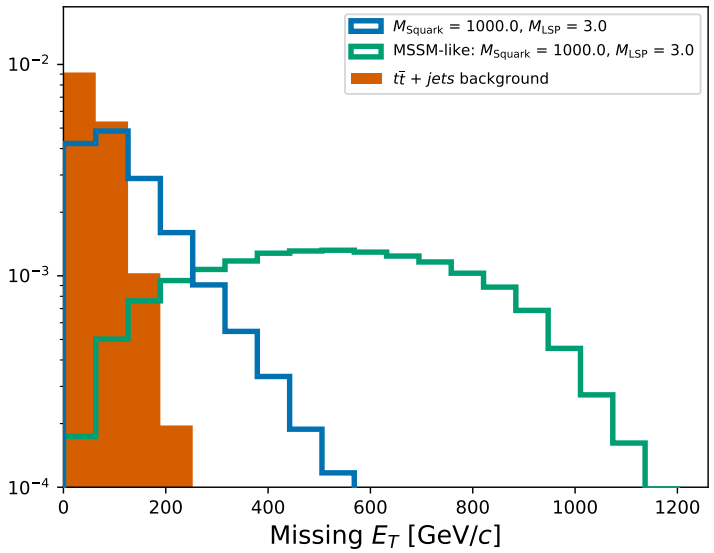


Figure 2.5: $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for the the BP1 benchmark point from Table 2.2 compared with that for $t\bar{t}$ MC background events and an MSSM-like simplified scenario.

Taking for example the first two BPs, denoted BP1 and BP2 in Table 2.2, we see that gluinos are in both cases around 10 GeV/c^2 heavier than squarks. Such a small mass gap was chosen in order to avoid the case where squarks much lighter than gluinos become unstable under radiative corrections [37]. The gluinos thus decay first into squarks, with each squark decaying into a NLSP and a correspondingly flavoured quark: $\tilde{g} \to \tilde{q} + q$; $\tilde{q} \to \tilde{\chi}_2^0 + q$. We see very low $E_{\mathrm{T}}^{\mathrm{miss}}$ spectra for both of these benchmark points, illustrated in Figure 2.4, with the average $E_{\mathrm{T}}^{\mathrm{miss}}$ slightly higher for BP2, with heavier squark and gluino masses. This is since the heavier squarks and gluinos result in a higher momentum NLSP, which in turn means that the LSP will also acquire higher momentum, albeit much less than the that inherited by the Higgs boson.

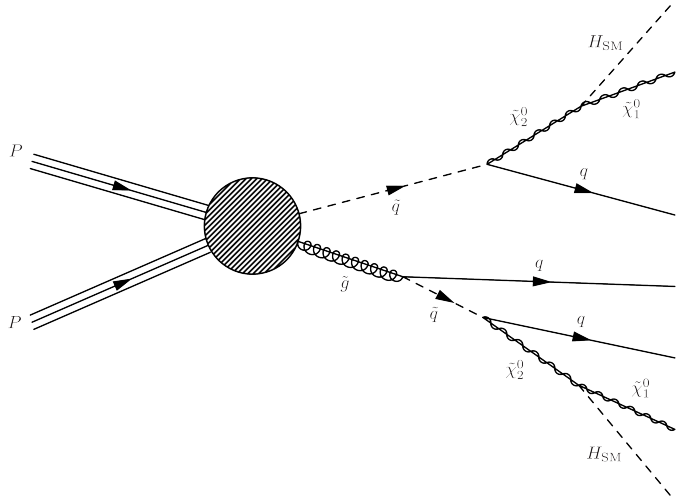
BP3 and BP4 differ in that the gluino is 200 GeV/c^2 lighter than the squarks. In these scenarios the left-handed squarks always decay into the gluino and a correspondingly flavoured quark, whilst the right-handed squarks decay either into a gluino-quark combination or skip this step entirely and decay directly into a NLSP and a quark, with corresponding BRs of 70% and 30%, respectively, with

these fractions close to those estimated by NMSDECAY [54, 55]. However, as shown in Figure 2.4, the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is very similar for these BPs compared with BP1 and BP2, with the heavier squark and gluino masses in BP4 resulting in higher average $E_{\mathrm{T}}^{\mathrm{miss}}$.

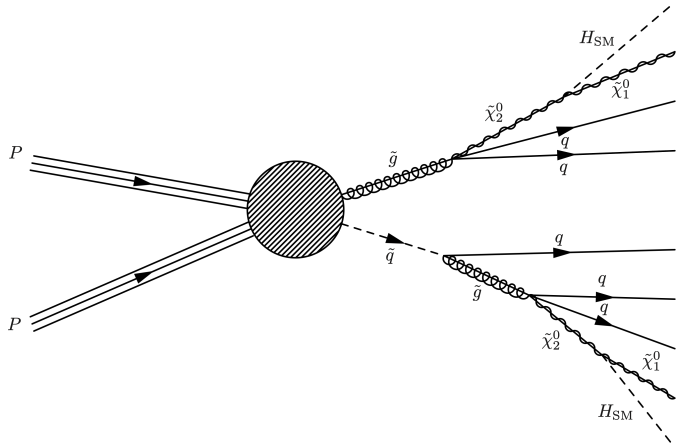
For BP5 and BP6 the squarks are lighter than the gluino, as in points BP1 and BP2. However, the respective stop/sbottom-type squark is now lighter than the gluino, sufficiently so that gluino two-body decays are possible. In these two BPs the gluino is assumed to always decay into a stop (BP5) or sbottom (BP6) squark and corresponding top/bottom quark, which in turn decays into a NLSP and correspondingly flavoured quark. The first- and second-generation squarks decay with 100% BR into the NLSP and corresponding quark as in BP1 and BP2.

BP7 and BP8 also involve the stop (BP7) or sbottom (BP8) squark, but with the gluino lighter than the first two generation squarks, as in BP3 and BP4. Much like in BP3 and BP4, the left-handed squarks always decay into a gluino and a correspondingly flavoured quark, whilst the right-handed squark decays with 70% and 30% BR into either a gluino and a quark or directly into a NLSP and quark, respectively. In both of these points the gluino always decays into the respective stop/sbottom squark and top/bottom quark, with the third generation squark decaying furthermore into a NLSP and corresponding top/bottom quark. In these final four BPs we see in Table 2.2 slightly higher average $E_{\mathrm{T}}^{\mathrm{miss}}$ values for those featuring stop squarks compared with the other BPs with similar squark and gluino masses. However the general $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution is still highly suppressed in all cases compared with that of a typical MSSM scenario, as illustrated in Figure 2.4.

Figure 2.6 shows example decay cascades for each of these BPs. BP2 and BP4 have been omitted, since the possible diagrams do not differ from those for BP1 and BP3, however, extra diagrams are included for BP3 and BP7 to illustrate the possible routes by which the right-handed squarks may decay. Additionally, the alternative BP7 decay chain may apply to BP8 by simply switching each of the stop/top for the corresponding sbottom/bottom squark/quark.

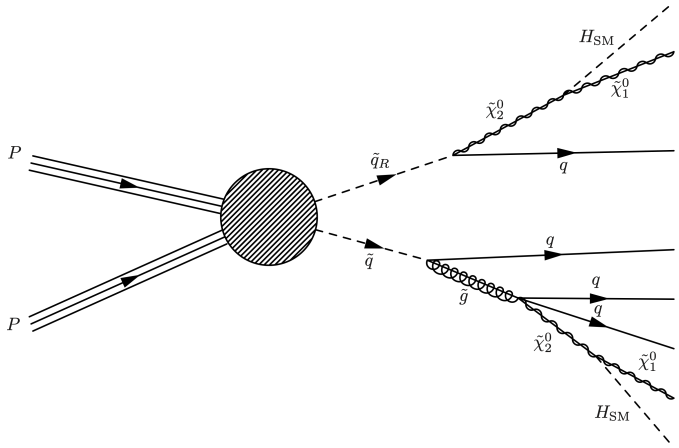


(a) BP1 possible decay cascade.

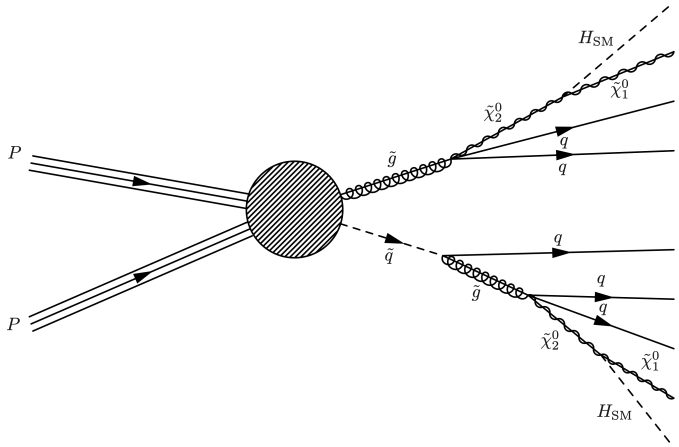


(b) BP3 possible decay cascade.

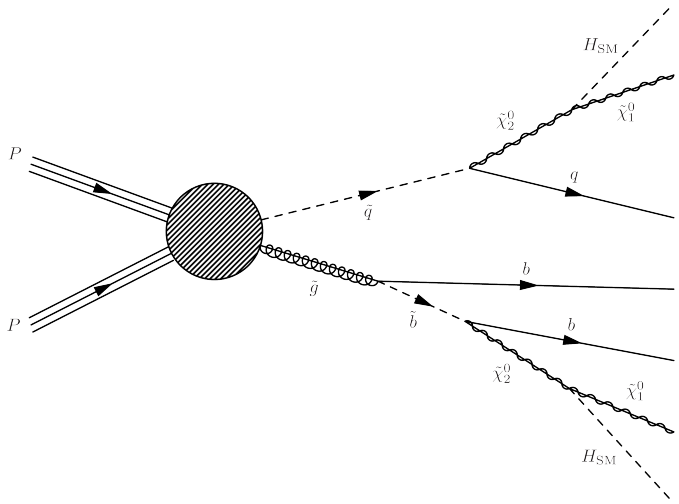
The Feynman diagrams in Figure 2.6 show examples of the processes by which we may produce a final state with two LSPs and two Higgs bosons. In each of these diagrams we produce one squark and one gluino directly, with each decaying via an example cascade, however, both squark-squark and gluino-gluino pair production processes are included additionally in the event generation. These processes may also involve extra hadronic jets produced along with the initial sparticles or typically softer radiated jets.



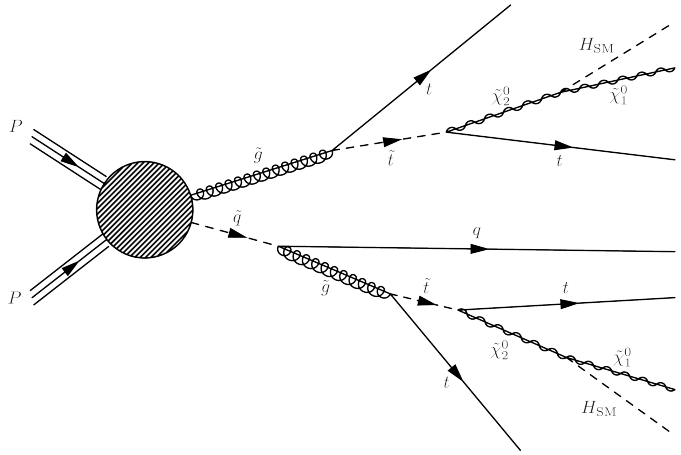
(c) BP3 alternative decay cascade.



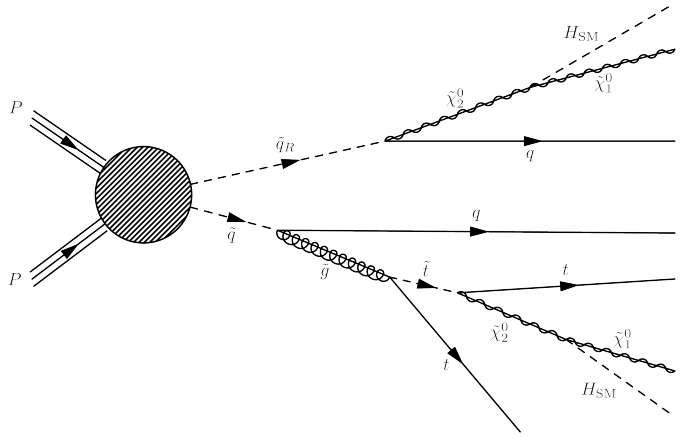
(d) BP5 possible decay cascade.



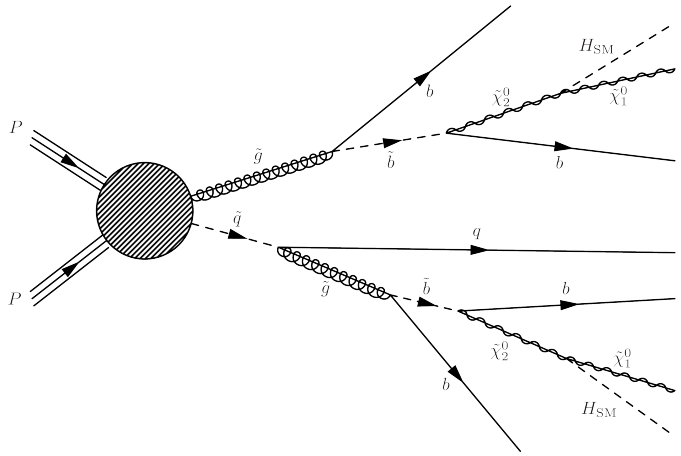
(e) BP6 possible decay cascade.



(f) BP7 possible decay cascade.



(g) BP7 alternative decay cascade.



(h) BP8 possible decay cascade.

Figure 2.6: Feynman diagrams showing example processes by which we may produce two singlino LSP along with two Standard Model-like Higgs bosons. An example diagram is given for each of the BPs in [37] which contain unique mass hierarchies. Here, BP1, BP3, BP5, BP6, BP7 and BP8 are shown in (a), (b,c), (d), (e), (f,g) and (h), respectively.

2.7 Summary

Throughout this chapter we have explored the theoretical motivations for supersymmetry and in particular the NMSSM, along with the possibility of scenarios within the NMSSM which feature highly suppressed $E_{\mathrm{T}}^{\mathrm{miss}}$ spectra. Example benchmark points have been introduced which exhibit small $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions, which would not be the case in the MSSM. In the next chapter we consider the LHC and the CMS experiment, data collected from which is used for the phenomenological and experimental analyses presented in this thesis.

3 | The Large Hadron Collider and the Compact Muon Solenoid Detector

3.1 The Large Hadron Collider

Located on the Franco-Swiss border near Geneva, the Large Hadron Collider (LHC) is a synchrotron designed to accelerate and collide protons, as well as heavy ions, up to a centre-of-mass energy of 14 TeV. With a circumference of 26.7 km, the LHC is the largest synchrotron in the world, and is situated on an oblique axis between 50 and 175 m underground, in order to account for geological factors during construction.

Four main experiments are based at the LHC: the Compact Muon Solenoid (CMS) [56], A Toroidal LHC ApparatuS (ATLAS) [57], LHC beauty (LHCb) [58] and A Large Ion Collider Experiment (ALICE) [59]. LHCb focuses on b-physics, with ALICE concentrating on the study of heavy ion collisions. ATLAS and CMS are both general-purpose detectors, and are situated as far apart as possible within the LHC ring in order to minimise any correlation in beam conditions between the two sites. The two latter experiments were responsible for the 2012 discovery of the Higgs boson [24, 25].

Firstly protons are obtained from a hydrogen source via ionisation within a strong electric field. These protons then undergo acceleration through LINAC2, a linear accelerator, reaching an energy of 50 MeV [60].

The protons are then accelerated to 25 GeV via the Proton Synchrotron (PS), as illustrated in Figure 3.1. Here the protons conglomerate into bunches of $\mathcal{O}(10^{11})$ protons, localised to within ~ 1 ns. The final acceleration prior to injection into the LHC is performed by the Super Proton Synchrotron (SPS), reaching an energy of

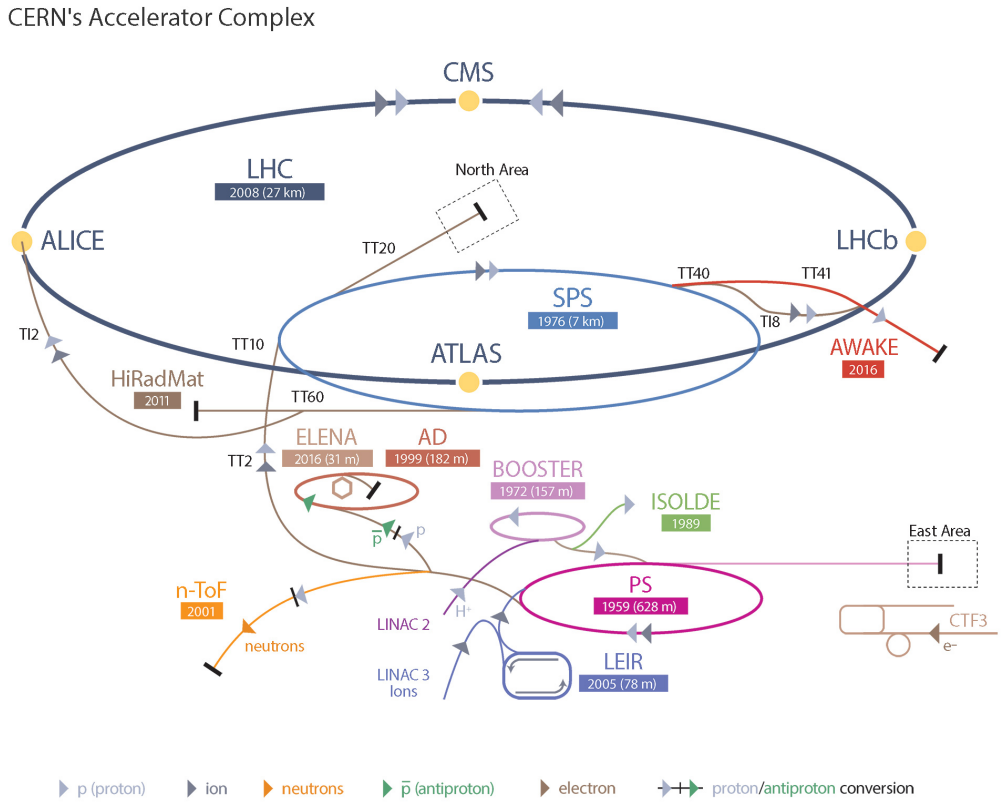


Figure 3.1: The CERN accelerator complex [61].

450 GeV. Proton beams are then fed into the LHC in opposite directions following energy adjustments.

Once circulating around the LHC, superconducting magnets cooled to 1.9 K with a current of up to 11.85 kA maintain the correct beam trajectory. Since the LHC accelerates same-charge protons, the two proton beams are kept in separate beam pipes, with equal and opposite magnetic fields.

Approximately 2076 bunches of $\mathcal{O}(11)$ protons circulate in the LHC during a normal pp collision fill, with bunches spaced by 25 ns; corresponding to a bunch crossing (BX) rate of 40 MHz. The actual rate of pp collisions is dependent on the instantaneous luminosity and inelastic cross-section. The luminosity is defined as follows:

$$\mathcal{L} = f \cdot \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}, \tag{3.1}$$

where f represents the frequency of BXs, n_1 and n_2 the number of protons in colliding bunches 1 and 2 respectively, and σ_x and σ_y respectively the root-mean-square horizontal and vertical beam sizes.

In environments where the instantaneous luminosity is very high, often more than one pp interaction occurs per BX. This phenomenon is referred to as in-time pileup (PU), since these interactions all stem from the same BX. Particles originating from a previous BX, often low-energy showers taking a longer time to leave the detector volume, are referred to as out-of-time pile-up (OOTPU). Both PU and OOTPU present challenges in reconstructing complex decay cascades, with a large effort involved in understanding and mitigating resultant issues.

3.2 The Compact Muon Solenoid Detector

The Compact Muon Solenoid detector is a general-purpose detector featuring a large magnetic solenoid, which causes the trajectories of charged particles to bend. The experiment uses pixel detectors, a silicon tracker, electromagnetic and hadronic calorimeters within the solenoid, and muon chambers outside the solenoid to reconstruct the traversed paths and decay vertices of particles produced in pp collisions.

CMS is located at Point 5 of the LHC, antipodal on the LHC ring to the ATLAS detector. The detector measures 21 metres long and 15 metres tall, consisting of 13 segments. Each segment was constructed independently on the surface before being lowered 95 metres underground. The largest segment, in the centre, contains the niobium titanium solenoid and had to be lowered over a course of 10 hours using a specialised hydraulic crane, in order to minimise any transverse movement and since the use of a conventional winch would have generated too much heat, causing the steel cables to fuse.

Despite its compact nature, being smaller than the other main LHC detectors, CMS is in fact the heaviest. The primary reasons for this is are the solenoid itself

and the magnetic field return yoke, formed of iron plates which sandwich the muon detectors outside the solenoid. Overall the mass of the CMS detector is around $12,500$ tons.

For illustrative purposes Figure 3.2 shows a 3D diagram of the CMS detector, taken from Ref [62]. Starting from the interaction point and moving outwards, the first part of the CMS detector we encounter is the pixel tracker, followed by the strip tracker. The next layers of CMS are the calorimeters: firstly the active lead tungstate crystal electromagnetic calorimeter (ECAL), with active referring to the crystal being a scintillating material, then the brass hadronic calorimeter (HCAL) and outer HCAL. The silicon trackers provide high-granularity particle position measurement close to the interaction point. ECAL and HCAL then measure energy deposits of electromagnetically and hadronically interacting particles respectively; albeit with coarser granularity in angular position.

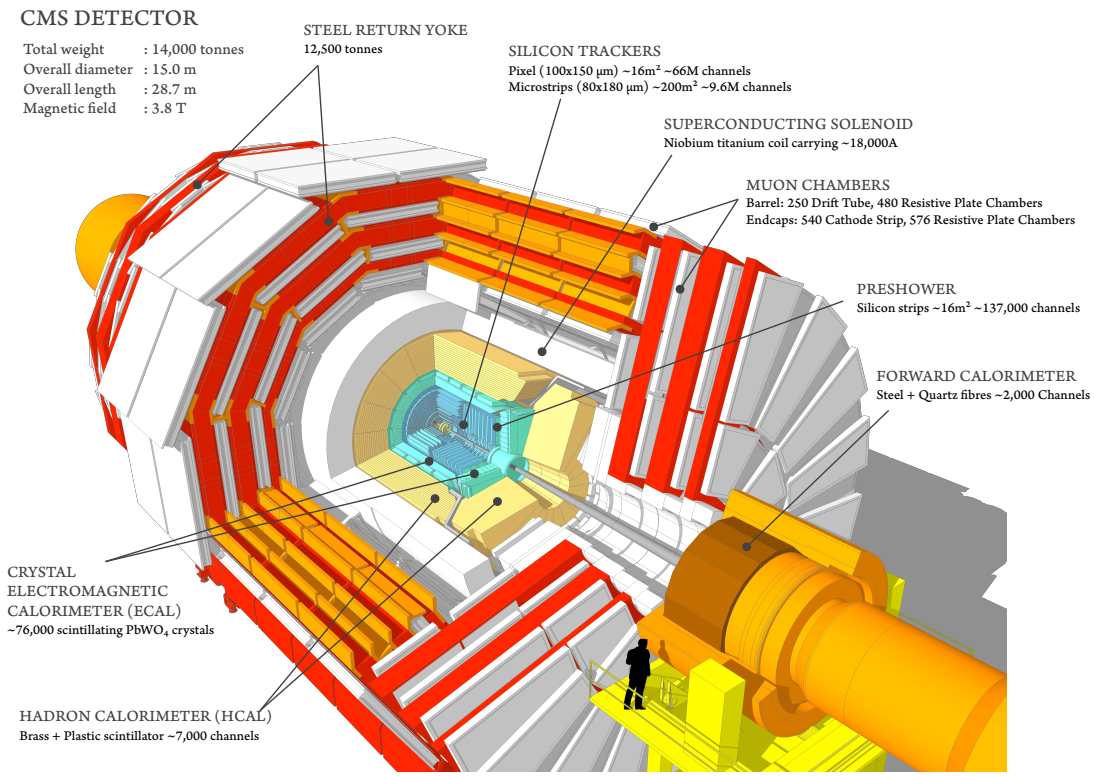


Figure 3.2: To-scale diagram of the CMS detector, taken from Ref [62].

At approximately half-radius is the niobium titanium superconducting solenoid, which generates an internal magnetic field of ~ 3.8 T parallel to the beam direction.

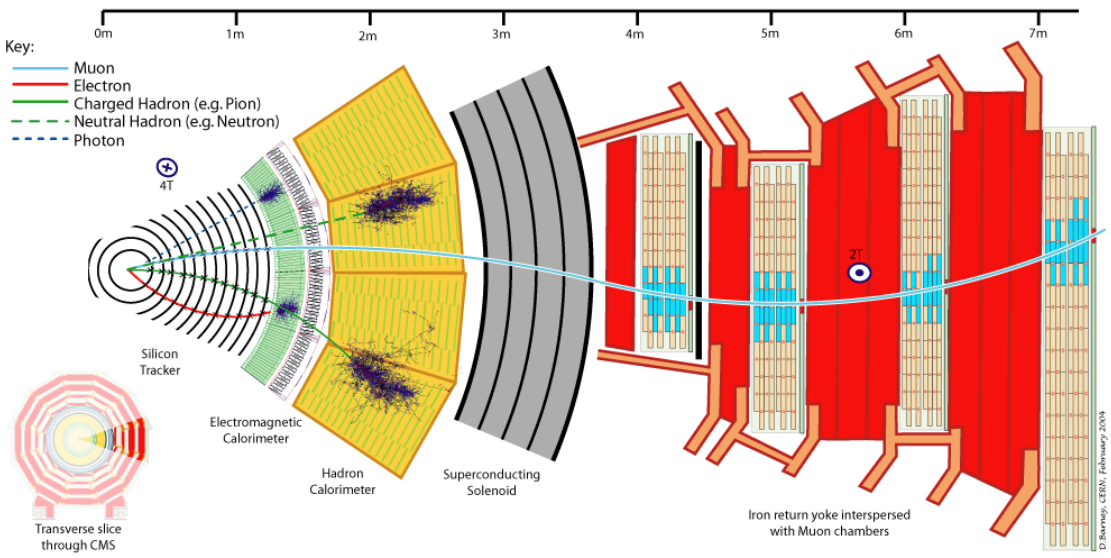


Figure 3.3: Slice in ϕ of the CMS detector, with example particle trajectories, taken from Ref [63].

Outside of the solenoid, between the iron return yoke plates, reside the three different types of muon detector. In this outer region within the return yoke the magnetic field is approximately ~ 2 T, again parallel to the beam direction but antiparallel to the field within the solenoid.

Figure 3.3 shows a slice in ϕ of the CMS detector, from the interaction point to the outer edge of the muon chambers. This diagram shows possible paths of example reconstructed particles. We note the curvature of the muon trajectory reverses around halfway out of the detector radius, since the magnetic field in the return yoke is antiparallel to that inside the solenoid.

3.2.1 Coordinate System

Geometrically the CMS detector uses a right-handed coordinate system. The x-axis points towards the centre of the LHC ring, the y-axis directly upwards and the z-axis along the beam direction. Particle position is described by two angles: azimuthal angle ϕ in the $x-y$ plane in the range $[-\pi, \pi]$ and pseudorapidity, η; a Lorentz-invariant quantity relating to θ, the angle from the z-axis.

This pseudorapidity, $|\eta|$, is defined as follows:

$$\eta := -\ln\left(\tan\left[\frac{\theta}{2}\right]\right), \tag{3.2}$$

where θ is the angle between the object in question and the z-axis.

The angular separation between two physics objects, e.g. particles or hadronic jets, is denoted ΔR. This is usually defined as:

$$\Delta R := \sqrt{\Delta\eta^2 + \Delta\phi^2}, \tag{3.3}$$

however in some cases the difference in object rapidity, $\Delta\theta$, is used in place of the pseudorapidity separation, $\Delta\eta$.

3.2.2 Silicon Tracker

The CMS tracker, 5.8 m in length and 2.5 m in diameter, is the closest part of the detector to the interaction point, whilst also being tasked with providing extremely high precision measurements of particle trajectories and secondary vertex detection; the latter being crucial for tagging objects such as bottom quark jets. As such, in order to provide precise angular resolution such a short distance from the primary vertex the tracker must have high spatial resolution. Between the barrel and endcap sections the tracker covers a pseudorapidity range of $|\eta| < 2.5$ [56]. The overall tracker resolution for a 100 GeV muon for example is $\sim 1 - 2$ % for the pseudorapidity range $|\eta| < 1.6$.

Furthermore, in order to be able to withstand the huge flux of particles passing through during periods of high instantaneous luminosity, as well as being able to resolve individual bunch crossings, the tracker must be radiation hard and have fast response. The estimated lifetime of the tracker is ten years.

In order to minimise effects such as multiple scattering, photon conversion and photon emission via bremsstrahlung, as little material is used in the tracker as possible. The inner tracker is based on layers of pixel detectors, with pixel size

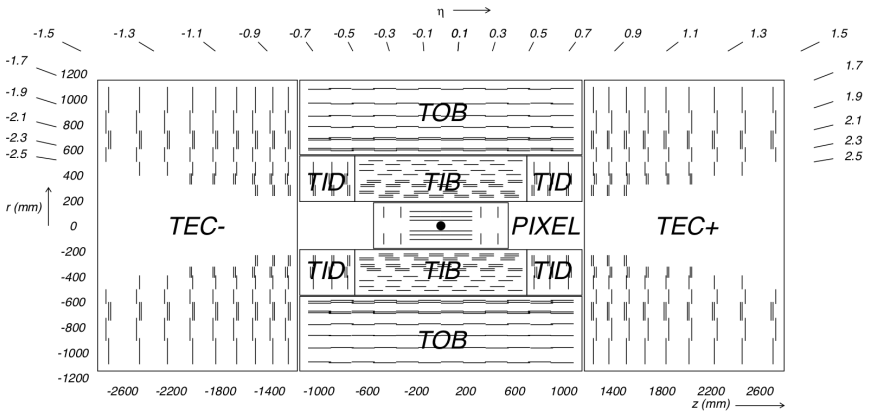


Figure 3.4: Illustration of the CMS tracker in the $r-z$ plane, from [56]. The pixel detector is shown closest to the interaction point, which is represented by a black dot.

of 100 x 150 µm in the $r-\phi$ and z directions; three layers in the barrel and two in each endcap, with the closest barrel pixel layer situated at a radius of 4.4 cm. The endcap pixel detector layers lie transverse to the beam axis, whilst the barrel layers lie parallel, as shown in Figure 3.4.

The silicon strip tracker resides in the volume covered by a radius between 20 cm and 116 cm and is formed 9.3 million strips of varying thickness depending on subsystem, with a combined active surface area of 198 m^2 [56]. The three key subsystems of the strip tracker are the Tracker Inner Barrel and Disks (TIB/TID), Tracker Outer Barrel (TOB) and Tracker Endcaps (TEC). Strips in the TIB and TOB have a width of 80 µm. Strips in the TOB have greater length than those in the TIB due to larger radius, 25 cm compared to 10 cm, but are required to be thicker – 500 µm compared to 320 µm – in order to combat the resultant increase in noise.

3.2.3 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is made up of lead tungstate ($PbWO_4$) crystals, designed to measure the energy of photons, electrons and positrons within the range $|\eta| < 3.0$. Within these highly dense crystals of size $\Delta\phi \times \Delta\eta = 0.017 \times 0.017$, of which over $75,000$ make up the ECAL volume, such particles quickly radiate energy due to the short radiation length, the length over which an electron

loses all but $\frac{1}{e}$ of its energy via bremsstrahlung, and a small Molière radius of 2.2 cm, the radius of a cylinder expected to contain 90% of the energy deposited by the electromagnetic shower from an incident electron [56]. Within the barrel region the ECAL crystals are 230 mm long, which corresponds to 25.8 radiation lengths. As electromagnetic particles radiate energy, light is scintillated within the ECAL crystals; this is then amplified by photodetectors [56].

Three subdetectors comprise the ECAL: the ECAL Barrel (EB), the ECAL Endcap (EE) and the ECAL Preshower (ES). In EB, covering the range $|\eta| < 1.479$, scintillated photons are collected using avalanche photodiodes (APDs). In the EE vacuum phototriodes (VPTs) cover the remaining range $1.479 < |\eta| < 3.0$, chosen due to their lower sensitivity to higher levels of radiation typically experienced in the forward regions.

Equation (3.4) represents the ECAL energy resolution, σ_E, as a function of particle energy, E:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 \oplus \left(\frac{N}{E}\right)^2 \oplus C^2, \tag{3.4}$$

where S represents a stochastic term for statistical fluctuations in the number of particles produced via scattering, N a term relating to the noise from the electronics and C a constant term. These terms were found to be $S = 0.028$ $(\text{GeV})^{\frac{1}{2}}$, $N = 0.12$ GeV and $C = 0.003$, resulting in an energy resolution of $\sim 0.5\%$ for a 100 GeV particle [64].

3.2.4 Hadronic Calorimeter

The second calorimeter in CMS is the Hadronic Calorimeter (HCAL), whose primary purpose is to measure the energy deposited by hadronically showering particles. The HCAL is formed of brass absorption layers alternated with plastic scintillator, with 17 layers making up the HCAL Barrel (HB) and 19 making up the HCAL Endcaps (HE). HB covers the range $|\eta| < 1.3$, with the HE at both ends covering the region $1.3 < |\eta| < 3.0$. A Forward HCAL (HF) extends this

range in the forward region up to $|\eta| < 5.2$, complementing HB and HE. Moreover, the HCAL is extended outside the magnet vacuum tank, embedded in the muon system, to ensure adequate sampling depth in the region $|\eta| < 1.3$. This extended section (HO) measures shower energy deposited after HB by using the solenoid coil as an absorber to promote interactions, increasing the depth in the barrel region to twelve interaction lengths, compared to ten in the endcap regions.

Nuclear interactions within the brass absorber plates create hadron showers, with the energy deposited causing scintillation in the plastic scintillator tiles. This frequency of this light is then reduced through wavelength shifting, and the light transferred along transparent optical fibres to hybrid photodetectors (HPDs). These HPDs then output an amplified response, corresponding to the original hadron energy.

In the highly-forward region the radiation conditions are much more harsh, and as such the HF uses a different technology compared with HB and HE. In this section steel absorber plates are embedded with scintillating quartz fibres, with charged particles in the shower emitting light which is then detected along these fibres.

Equation (3.5) shows the energy resolution σ_H of HCAL for a particle of energy E [65].

$$\frac{\sigma_H}{E} = \frac{0.943(\text{GeV})^{\frac{1}{2}}}{\sqrt{E}} \oplus 0.084. \tag{3.5}$$

3.2.5 Superconducting Solenoid

A key feature of the CMS detector is the superconducting solenoid, which generates a 3.8 T homogeneous magnetic field parallel to the beam axis, sufficient to bend the trajectories of charged particles for the purposes of particle identification and momentum reconstruction [56]. Five aluminium modules comprise the solenoid,

each 2.5 m long, 6.3 m in diameter and containing four winding layers of reinforced, stabilised Niobium-Titanium (NbTi) superconductor.

Outside the solenoid radius an iron return yoke contains the majority of the magnetic field, with this field measuring approximately 2 T, antiparallel to the field within the solenoid. This yoke comprises five barrel wheels and six endcap disks, weighing around $10,000$ tons; most of the mass of the entire detector. A vacuum cryostat isolates and cools the magnet using liquid helium.

3.2.6 Muon Detectors

Due to their much higher mass, muons are less affected by the Coulomb force when passing near nuclei in a material. Therefore muons are able to pass much further through dense material compared with electrons of similar momenta, and typically pass through the CMS detector, leaving a clean signature. Precise momentum and position reconstruction is possible with the CMS muon system, a key component of the CMS detector consisting of three subsystems: the Drift Tubes (DT), the Resistive Plate Chambers (RPC) and the Cathode Strip Chambers (CSC), with each interspersed between the iron return yoke plates. Figure 3.5 illustrates the relative positions of the three respective components of the muon system.

Between the DT, RPC and CSC subsystems the muon chambers cover the range $|\eta| < 2.4$. DT detectors lie parallel to the beam axis in the barrel region, for which $|\eta| < 1.2$, where the magnetic field is uniform. The endcap region, $0.9 < |\eta| < 2.4$, contains CSC detectors due to their fine segmentation, since this region contains high muon flux and the magnetic field is non-uniform. These strips are arranged in wheels, with the cathode strips aligned radially, with perpendicular anode strips. In addition, dual-layer RPCs are employed in the barrel and endcap regions for which $|\eta| < 2.1$. These allow for the correct muon tracks to be more reliably distinguished in the case where multiple hits in the muon chambers cause the formation of ambiguous tracks.

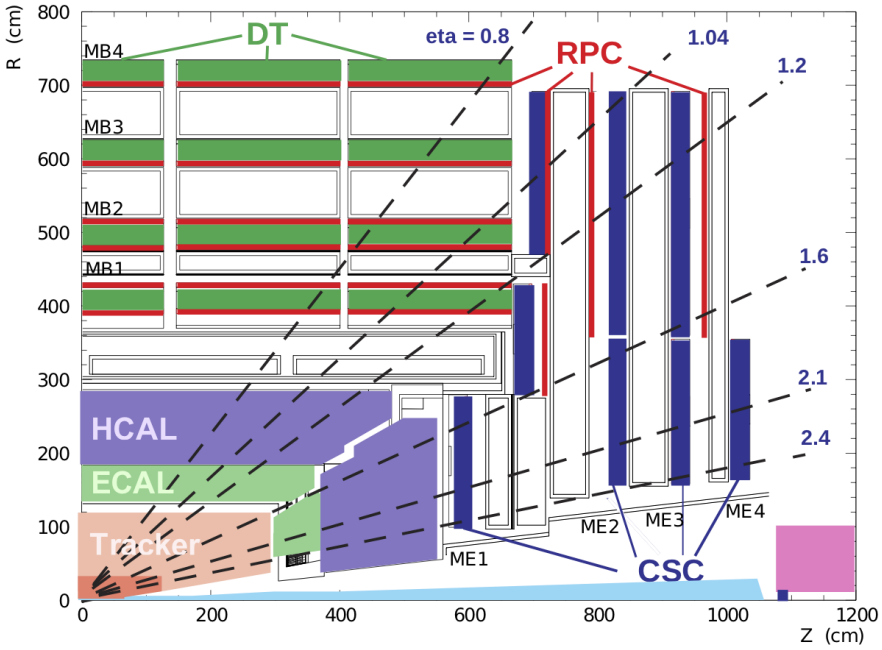


Figure 3.5: Illustration of the CMS muon system, featuring the DT, RPC and CSC subsystems, from [66].

3.2.7 Trigger

The rate of data generated per bunch crossing at an LHC experiment is approximately 1 MB. Given the bunch crossing rate of 40 MHz the raw, unfiltered data rate output from the CMS detector readout is around 40 TB of data per second, amounting to more data produced in one hour than Facebook collects in one month [67].

Writing the entirety of this data stream to disk would not only be a monumental task, but also an unnecessary one. The majority of this data stems from low-energy scattering events, often producing no final states of interest when searching for new physics at high energies. Owing to the Data Acquisition System (DAQ) bandwidth and the sustainable speed at which data streams may be written to tape therefore, not all data is recorded.

In order to decide efficiently which bunch crossing events to record to disk, a trigger system is employed, accepting or rejecting data depending on the output of a collection of algorithms. The CMS event trigger is formed of two systems: the Level-1 trigger and the High Level Trigger (HLT).

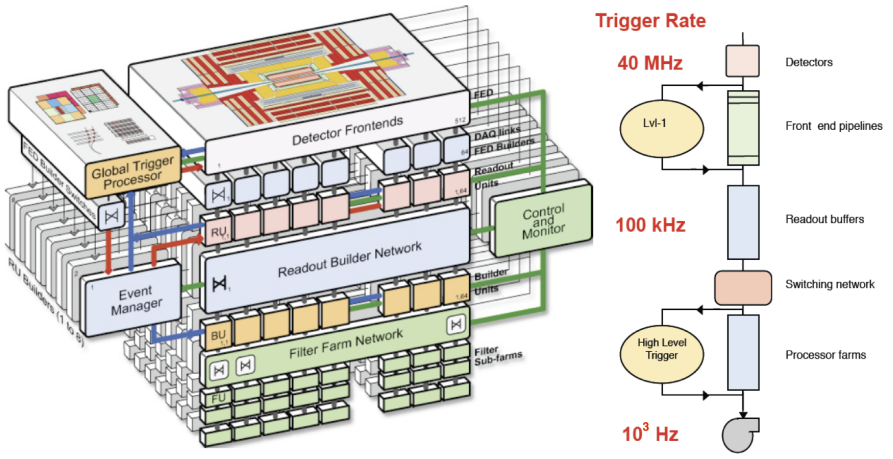


Figure 3.6: Illustration of the CMS DAQ system, including the input and output rate of the two trigger systems, taken from [68].

Figure 3.6, taken from Ref [68], shows an illustration of the CMS DAQ architecture, from the initial detector readout rate of 40 MHz down to the final 1 kHz which is written to disk.

Level-1 Trigger

The Level-1 (L1) trigger is based on Field Programmable Gate Arrays (FPGAs), Application-Specific Integrated Circuits (ASICs) and custom electronics situated underground in the CMS service cavern, close to the experiment itself, and is the first stage of data analysis. The design of the L1 trigger is to reduce the event rate from the initial 40 MHz to 100 kHz, before the remaining events are considered by the HLT.

The L1 trigger considers information from the ECAL, HCAL, HF and muon chambers every 25 ns in order to reject an event or pass it to HLT. However, due to hardware constraints the firing of the L1 trigger means the next two BXs are missed. The readout from these subdetectors is then used by the muon trigger to reconstruct muons, and by the calorimeter trigger to reconstruct electrons/photons (e/γ), tau leptons and jets, as well as quantities relating to energy sums, such as H_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$.

The full detector readout is stored in memory awaiting the L1 trigger decision, and can be held for a maximum of 3.2 µs. In order to ensure the L1 trigger is

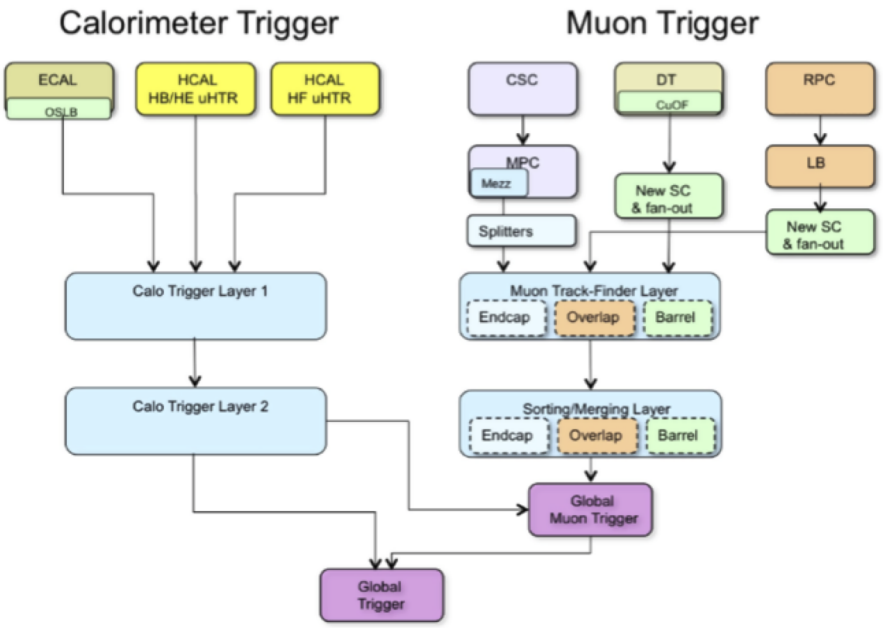


Figure 3.7: Illustration of the CMS upgraded trigger architecture, featuring the calorimeter trigger (left) and muon trigger (right), from [69].

able to read the data, perform its algorithms and send the output to the front-end electronics, the granularity of the calorimeter and muon detector information used is quite coarse, with the algorithms being simple compared with the equivalent offline reconstruction algorithms. Furthermore the tracker information is not currently used in the L1 trigger due to its enormous complexity, though future trigger upgrades do include such integration.

For Run-II of the LHC the CMS L1 trigger consists of the Global Trigger (GT), Global Muon Trigger (GMT) and Calorimeter Layer-1 and Layer-2 triggers [69]. Figure 3.7 illustrates the flow of data from the initial detector readout to the GT.

Physics objects reconstructed by the calorimeter and muon triggers are passed on to the GT. Depending on whether any of these objects pass the various event selection criteria, such as E_{T} thresholds and angular separation, the full detector readout is then passed to the HLT. Should the L1 trigger rate be too low, the data collection rate will be sub-optimal. Too high however, and the maximum allowed bandwidth will be exceeded, overloading the L1 trigger hardware and causing deadtime; whereby BXs are missed due to the trigger not being ready to record data.

Each trigger "bit", the true/false bit representing whether a particular trigger criterion has been passed, has a *pre-scale* factor. This factor allows for a fraction, the inverse of the pre-scale factor, of the events passing a given selection criterion to be passed to the HLT. Triggers which many events are expected to pass are usually reduced by a large pre-scale factor, with high-threshold triggers given a pre-scale factor closer to unity. The various trigger pre-scale factors may then be chosen carefully in order to keep the L1 trigger acceptance rate close to, but not exceeding, 100 kHz, and reduced as the instantaneous luminosity decreases during beam collisions as fewer and fewer proton bunches remain in the LHC.

Upgraded Calorimeter Trigger

An upgraded L1 trigger was used for Run-II of the LHC, reaching full operational status by the beginning of run year 2016. The primary reason for the upgrade was that an increase in instantaneous luminosity beyond the design specification of CMS was foreseen, with typical values in run-year 2018 being around 1.9×10^{34} cm$^{-2}s^{-1}$ – almost twice the design specification of 1.0×10^{34} cm$^{-2}s^{-1}$. Furthermore, the expected number of pileup interactions per BX increased from the design specification of 25, with run-year 2018 featuring an average of 37. In order to address these increases and to maintain good trigger performance with an output rate of less than 100 kHz, the calorimeter and muon triggers were upgraded.

The calorimeter trigger takes inputs known as trigger primitives, originating from ECAL and HCAL. Each trigger primitive relates to a respective sector in (η, ϕ) within the detector, known as a trigger tower. The ECAL trigger primitives store the E_{T} deposited in each respective ECAL trigger towers as an 8-bit integer, along with a flag indicating whether the deposit is best compatible with an electron or a photon. Similarly the HCAL trigger primitives are also stored as 8-bit integers, with a flag indicating deposits originating from minimum ionising particles. The angular size of the trigger towers is uniform in the barrel region of the calorimeters, increasing in $\Delta\eta$ as $|\eta|$ increases beyond 1.740 to align with the position of the physical HCAL towers.

Following the upgrades to the L1 trigger, the calorimeter trigger is formed of two layers. The two calorimeter trigger layers form part of the Time Multiplexed Trigger (TMT), also introduced as part of the L1 trigger upgrade [70]. The purpose of the TMT is to allow for the physics object reconstruction algorithms to be executed on a single processing node, using all of the ECAL/HCAL trigger primitive data for a given BX, rather than splitting the data for one BX across many processing nodes.

In order to accomplish this the role of the Layer-1 hardware is to process the ECAL and HCAL trigger primitives for a given BX. Layer-1 uses CTP7 cards, with the previously used copper links to ECAL having been replaced with optical fibres. The information for one BX is then sent to a single Layer-2 MP7 processing card. A different MP7 receives the information corresponding to the next BX and so on, until none remain and so the first MP7 is used once more. This allows each MP7 far longer than 25 ns in which to receive the entire trigger primitive data and execute the physics object reconstruction algorithms. Optical fibre links connect the 18 Layer-1 CTP7 cards to the Layer-2 MP7 cards; most of which are active, with redundant nodes available in the case of card failure, as illustrated in Figure 3.8.

High Level Trigger

The HLT is next in the data pipeline, and is formed of a large array of computers. The HLT performs higher precision calculations using more sophisticated algorithms, which are able to take advantage of the greater decision time allowed, since only events accepted by the L1 trigger are analysed by HLT. The HLT reduces the event rate of $\sim$ 100 kHz from the L1 trigger down to a rate of $\sim$ 1000 Hz, the maximum sustainable rate at which event data may be written to disk.

The HLT uses the information passed from the L1 trigger to identify physics objects, in order to reconstruct these objects more efficiently. Objects are then reconstructed in stages, with each stage becoming more complex. The final stage utilises the tracker information, which is not present in the L1 trigger. The final

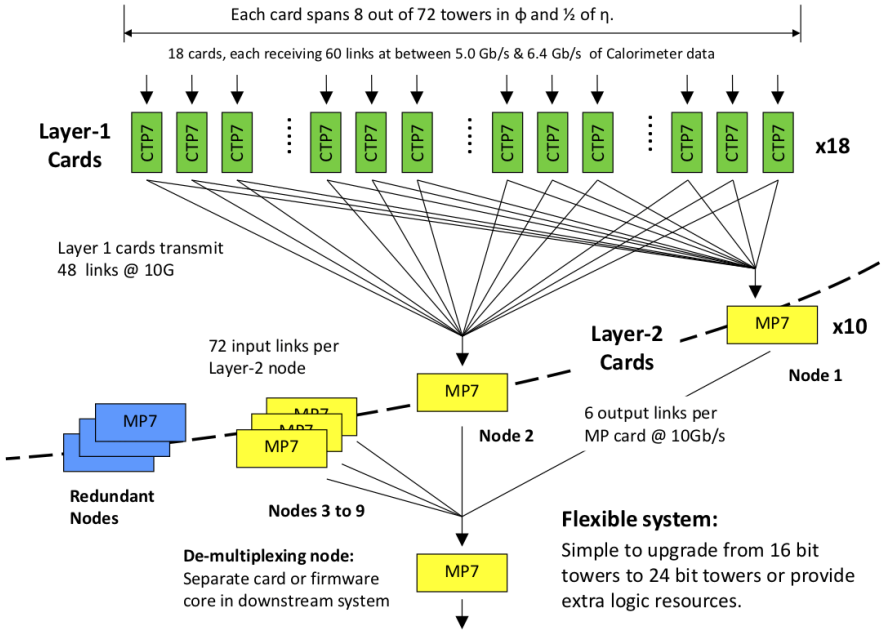


Figure 3.8: Illustration of the CMS upgraded calorimeter trigger, featuring the Layer-1 and Layer-2 subsystems, from [70].

objects reconstructed by HLT are then compatible with those reconstructed using full offline reconstruction.

3.3 Summary

As we have seen in this chapter, the CMS detector at the LHC allows for the detection and measurement of particles produced from pp collisions, with the various sub-detector components allowing for the precise reconstruction of different particle species. In the following chapter we will examine the software tools used in order to interpret the data collected at the CMS detector in the context of reconstructed particles, as well as those used to simulate the response of such a detector to MC generated events corresponding to novel signal processes and existing SM background processes.

4 | Software and Simulation Tools

In order to search for new physics at a collider such as the LHC we must first model the signal and background processes, in order to explore how each process would appear within a detector at a given centre-of-mass energy. Monte Carlo (MC) simulation techniques allow for the generation of large numbers of individual particle physics events which behave according to the probability density functions (PDFs) of a given model. These PDFs may be developed using purely theoretical methods, from data-driven approaches, or a combination thereof.

Event simulation begins with an event generation stage (GEN), where the hard scattering matrix elements are calculated for the initial state partons and events generated accordingly. During this stage the respective decays of the particles are performed, along with showering and hadronisation. From this stage onwards the method differs slightly depending on whether an experimental or phenomenological approach is taken: for an experimental analysis the GEN stage is followed by simulation (SIM), digitisation (DIGI) and reconstruction (RECO).

During the simulation stage the generated particles' interactions with and paths through the detector are computed, with the digitisation stage simulating the detector response to the resultant energy deposits. Finally the reconstruction stage outputs collections of physics objects akin to those reconstructed in the actual detector when taking data from real collision events.

For a phenomenological approach the main philosophy is unchanged, however the detector modelling is less precise. The GEN stage is largely the same, however the SIM, DIGI and RECO stages are combined and a parameterised approach is taken, approximating quantities including flavour tagging in quarks, hadronic jet energy resolution depending on properties such as particle flavour and momentum, as well as those of neighbouring particles. This approach not only reduces

runtime but also allows for theorists and phenomenologists to explore how a given theoretical model might appear within a particular detector without requiring a full-scale experimental analysis to be undertaken within a large collaboration, since collaboration software is not usually publicly available for use.

4.1 Event Generation, Showering and Hadronisation

The event generation and non-perturbative showering/hadronisation steps represent the GEN stage. In this section the software tools used to perform these steps are described, with any differences in these processes between phenomenological and experimental applications detailed.

4.1.1 MadGraph

For the initial matrix element-level hard scattering processes MadGraph 5 v2.3.3 [71] is used to generate the signal processes at leading order (LO). Other event generators are available, such as CalcHEP [72], POWHEG [73], Herwig [74] and SHERPA [75], each with its own benefits and intended use cases. CalcHEP is based around calculating the sum over gamma matrices, rather than using the individual particle helicity as is the case in MadGraph. POWHEG on the other hand combines NLO matrix element-level and parton shower calculations to give accurate results for particles produced at high and low p_{T} values. However, unlike MadGraph and CalcHEP, POWHEG does not consider NMSSM processes and so is not appropriate for the purposes of generating simulated MC events representing the NMSSM scenarios considered in this thesis.

Following its use in the theoretical work in Ref [37], from which mass scans are developed in this thesis, along with CMS SUSY searches such as Ref [15] which is considered in Chapter 6, we opt for consistency to use MadGraph for all parton-level signal model event generation in this thesis.

Many of the background process simulated events produced by the CMS collaboration for use in experimental analyses are generated to next-to-leading order (NLO), in order to better model final state production mechanisms which have non-negligible contributions stemming from Feynman diagrams containing loops. These Monte Carlo background events are usually produced by a dedicated group within the collaboration, due to the complexities involved and the large amount of time taken to generate huge numbers of events at NLO.

However this is not currently possible for the signal processes considered, as NMSSM squark and gluino production at NLO is not yet implemented directly within event generators such as MadGraph. In order to account for corrections from one-loop diagrams the total squark-squark, squark-antisquark, (anti)squark-gluino and gluino-gluino production cross-sections are calculated at NLO using Prospino, detailed in Section 6.2. Furthermore the emission of up to two hard jets is allowed in MadGraph.

4.1.2 Pythia8

Pythia8 [76, 77], a complete C++ rewrite of the Fortran-based Pythia6 (following the short-lived Pythia7 project), is a software package used in this work for performing the resonance decays and hadronisation of particles within simulated events.

The events stored within the output `.lhe` file from MadGraph contain properties such as the partons' mass, particle type, flavour and momenta. These, along with model- and mass point-specific features such as mass spectrum, particle decay branching ratios and various mixing matrices, are used as input to Pythia8 in order to determine possible decay routes and calculate final state particle momenta.

Furthermore Pythia8 includes the radiation of quarks and gluons from partons originating from MadGraph. Particle hadronisation, whereby individual partons form observable hadrons, is also performed by Pythia8, using the Lund string fragmentation framework [78]. Another approach is the cluster model [79], as used

in tools such as Herwig [74] and SHERPA [75]. Both models describe data from colliders well, the cluster model slightly less so than the string model but whilst containing fewer parameters, whose values are derived from data [80].

4.2 Jet Matching and Double Counting

4.2.1 Jet Matching

When interfacing a parton-level event generator such as MadGraph with Pythia for the purposes of showering and hadronisation one must take into account the matching of hadronic jets between generators. For the simulation of quarks and gluons produced via high-energy parton interactions Madgraph is used, since the MC techniques employed by Pythia break down in the cases of high-momentum and widely separated hadronic jets [77]. In the case of softer, radiated jets however, Pythia is used, since the matrix element-level calculations in MadGraph begin to diverge as the partons become soft or collinear [71, 77]. In order to correctly merge the MadGraph parton-level events with those in Pythia after parton showering has been performed, MLM merging with the following user-selected jet matching thresholds is employed: for MadGraph we have `XQCut` and for Pythia `QCut`.

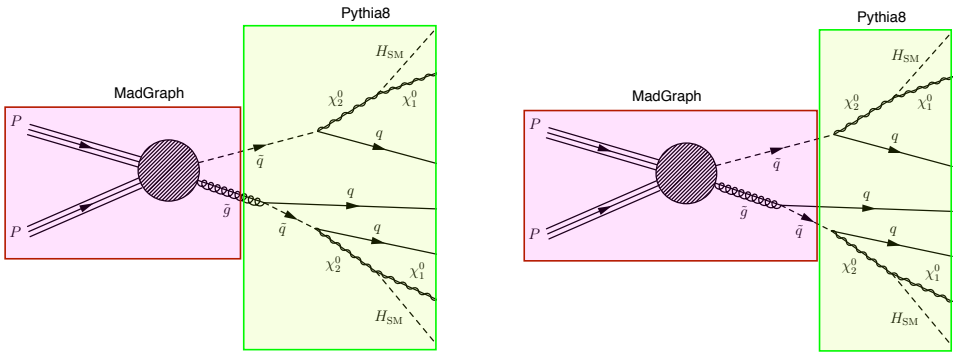
Firstly, we define a variable k_T between two partons, i and j, as follows:

$$k_T = \sqrt{2 \cdot \min(p_{Ti}, p_{Tj}) \cdot [\cosh(\eta_i - \eta_j) - \cos(\phi_i - \phi_j)]}, \tag{4.1}$$

where in the case of only one hadronic jet k_T is equal to the jet p_{T}.

In the case in MadGraph where $k_T <$ `XQCut`, meaning the partons are not well enough separated, the event is discarded. Hadronic jets whose p_{T} exceeds this threshold are kept from MadGraph, whereas others are discarded in favour of low-p_{T} jets generated in Pythia.

Moving to Pythia, the k_T is calculated between every final state particle. These are then combined until all remaining pass the `QCut` threshold. If this is too high,



(a) Gluino decays in Pythia.

(b) Gluino decays in MadGraph.

Figure 4.1: MadGraph event generation of two squarks and one jet (left) and one squark and one gluino (right), sharing the same Feynman diagram after Pythia has performed the gluino decay.

not enough jets will pass the threshold and so the event is vetoed, avoiding the undesirable case where Pythia produces high-p_{T} jets. Too low however, and the event is also vetoed, avoiding the equally undesirable case where Pythia would not be allowed to produce low-p_{T} jets. We nevertheless lose a number of events at the jet matching stage, since events are vetoed such that no double-counting occurs.

4.2.2 Avoiding Event Double Counting

Considering the Feynman diagram in Figure 4.1, an initial gluino and squark may be produced at parton-level in MadGraph, with the second squark produced by Pythia from the decay of the gluino, as demonstrated in figure 4.1a. However, if one were to direct MadGraph to produce two squarks, allowing the production of extra jets, some of the sub-processes considered would involve a squark stemming from the decay of a gluino, all of which would be calculated at parton-level, as shown in Figure 4.1b.

Thus there is an overlap between the events generated by MadGraph when asked for two squarks and those generated when producing gluinos, which are in turn decayed by Pythia into squarks.

In order to remove the possibility of sub-processes being counted more than once in the overall calculation, it is required that any squark or gluino whose decay is performed at parton-level must be off-shell, with no resonance allowed. Since any

squarks or gluinos whose decay is performed by Pythia will have been treated as *final state* particles within MadGraph, they must be on-shell. Thus, by summing the complementary on- and off-shell terms, the entire momentum space over which the squarks and gluinos may decay is correctly obtained.

4.3 Detector Simulation: Delphes

One tool for detector simulation in phenomenological work is Delphes [81]. Rather than precisely modelling every part of a detector and emulating its response to particles in simulated events, the detector behaviour is instead parameterised. When using Delphes for the purposes of work presented in this thesis, any experiment-specific parameters are taken to be those provided in the CMS Delphes configuration card [81].

4.3.1 Tracker

Firstly the tracker response is calculated by Delphes, depending on the particle's charge. The magnetic field is assumed to be uniform and localised within the inner tracker region, and particles originating outside this region are ignored. Depending on the charge of a particle, it follows either a straight or helicoidal path until it hits the calorimeters.

Charged particles are reconstructed as tracks in the tracker volume with a certain probability, dictated by experiment-specific user-defined parameters. Smearing is applied to the magnitude of the transverse momentum vector of tracks, but perfect angular resolution is assumed. Energy and momentum resolutions are also specified in the Delphes input parameters, with dependence on particle type, pseudo-rapidity and transverse momentum.

4.3.2 Calorimeters

The electromagnetic and hadronic calorimeters, ECAL and HCAL respectively, follow after the tracker. Each of the calorimeters is segmented into towers, with

the granularity assumed uniform in ϕ and equal between ECAL and HCAL in ϕ and η directions, for the sake of minimising computation time, with the calorimeter energy deposit calculated as the geometric centre of a given calorimeter tower.

By default all electrons, positrons and photons deposit 100% of their energy into ECAL. Similarly all hadronically decaying particles' energy is deposited into HCAL. This therefore differs from a real detector, where it would be expected for some hadronic particle energy to be deposited into ECAL, as well as for very highly energetic electromagnetic particles to shower beyond the outer edge of ECAL and into HCAL – a phenomenon known as *punch-through*.

The ECAL and HCAL resolutions are parameterised as functions of the energy and pseudorapidity of the particle, as follows:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S(\eta)}{\sqrt{E}}\right)^2 + \left(\frac{S(\eta)}{\sqrt{E}}\right)^2 + C(\eta)^2, \tag{4.2}$$

where σ differs between ECAL and HCAL and S, N and C are stochastic, noise and constant terms respectively [81].

A validation exercise is carried out in Section 6.4.7 in order to compare the performance of Delphes compared with proprietary CMS detector simulation software, which is detailed below.

4.4 Detector Simulation: CMSSW

CMS offline software (CMSSW) is a collection built around an event data model (EDM). Its construction is modular, allowing for plugins relating to simulation, calibration, alignment and reconstruction. This software collection forms the basis upon which all experimental analysis code used in this thesis are built.

The EDM means that all data stored is on an event-by-event level. The running of CMSSW centres around the main executable `cmsRun`, where the parameters such as which data to use, which modules to include in what order are set at

runtime. This means the same executable handles both MC and data events, and handles the processing of events from the DIGI stage through to RECO and from there into ntuples; collections of analysis-specific event variables with any necessary calibrations and corrections applied.

4.4.1 Full Event Simulation in CMSSW

In order to perform full detector simulation in CMSSW for the generated signal mass points used in this thesis, we begin with the `.lhe` files containing parton-level events generated using MadGraph. Similarly to the case for the phenomenological studies in Chapter 6, Pythia8 is used to perform the showering and hadronisation step, run within CMSSW.

Following the GEN stage, the SIM stage simulates the interactions between the simulated particles and the detector material using Geant4, a comprehensive toolkit used for simulating particles travelling through and interacting with matter [82–84]. Therefore as opposed to Delphes, which applies parameterised resolution and smearing functions when simulating the detection of objects in an event, the path traversed by each object is simulated fully, along with its interactions with the various materials in the detector.

After this stage the DIGI stage simulates the electronic response to the energy deposits left in the various detector volumes. This DIGI data contains the full detector information per event, including every detector hit, every measurement in every calorimeter channel, and the pass/fail bits for every trigger. This is then processed into RECO data by performing the event reconstruction according to the methods detailed in Section 5.1. This RECO data format stores information such as reconstructed particles with any associated variables, such as those used in calculating object isolation, as well as reconstructed tracks.

Finally, the RECO data is converted once more into miniAOD, a compressed form of the AOD (Analysis Object Data) format. This format is similar to RECO, but drops information such as individual calorimeter hits in order to reduce file size.

In the case of simulated data, events are stored in the miniAODSIM format, which is essentially the same, but in addition to the miniAOD contents this also contains the generator-level objects, describing the simulated partons before any detector simulation or smearing is applied. These objects are often useful, for example in validating reconstruction methods or for deriving re-weighting factors for a given simulated object.

4.5 Summary

In this chapter we have examined the software tools which are used as part of the phenomenological and experimental analysis works presented in this thesis. The work shown in Chapter 6 uses only publicly available data and software, for which the Delphes detector simulation framework is used. For the CMS experimental analysis detailed in Chapter 7 the full CMSSW framework is used, in order to allow for the most accurate simulation of the expected detector response to the considered NMSSM signal and SM background processes, as well as precise reconstruction of real pp collision events in data. In the following chapter we will examine the measurement variables which will be used in these respective analyses, the calculation of which depends on the performance of the software tools explored in this chapter.

5 | Objects and Measurement Variables

This chapter defines a number of key kinematic variables used as part of the analyses considered in this thesis, along with the criteria observed when reconstructing and identifying various particles. Having considered the methods by which various objects are reconstructed in Chapter 5, as well as the approximations thereof using Delphes, we now look at more analysis-specific definitions and reconstruction criteria.

In the case of the phenomenological study in chapter 6 we examine the relevant variables used in [15] as well as the criteria for inclusion of objects such as hadronic jets and isolated leptons. For the CMS analysis detailed in chapter 7 the remaining objects and variables are introduced, in addition to any minor differences in object reconstruction criteria compared with those used in the previous phenomenological work.

Certain reconstructed objects such as hadronic jets appear throughout the various works within this thesis, both experimental and phenomenological in nature. Whilst the base concepts remain the same, there are sometimes minor differences in the calculation processes, which will be detailed in this chapter.

5.1 CMS Object Reconstruction

Having acquired detector readout data from proton collisions, as described in Section 3.2, these readouts must be combined in order to ascertain the types of particles in the shower and their properties. Data from a bunch crossing is collected into an event, with the readouts for each event translated into physics objects. These physics objects have precise definitions, along with corrections and calibrations, which are provided by the CMS collaboration.

5.1.1 Particle Flow Reconstruction

Since particles will often pass through more than one sub-detector volume, it is advantageous to combine the readouts of these respective channels. The Particle Flow (PF) algorithm [85] considers information from all channels of the detector, providing better particle-based reconstruction than could be possible with only one sub-detector.

Not all particles have long enough lifetimes or strong enough interactions to measure directly in a detector. The main particle species which feature in CMS measurements are $e^{\pm}$, γ, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, K^{0}, $p^{\pm}$ and n, with these being categorised into the following groups by the PF algorithm: electrons, photons, muons, neutral hadrons and charged hadrons.

Starting from collections of tracks, calorimeter deposits and other sub-detector objects, the easiest particles to reconstruct without ambiguity are considered first, and are removed from the collections. This allows for the more difficult particles to reconstruct, neutral hadrons, to be considered last, by which point the removal of other detector hits should place constraints on any remaining particles, aiding reconstruction.

PF muons are reconstructed first, with muons identified in the muon chambers matched with hits in the tracker. These tracks and muon detector hits are then removed from the respective collections, with the deposited muon energy subtracted from any HCAL and ECAL regions passed through, since muons are minimally ionising particles (MIPs) and therefore leave small energy deposits whilst passing through the calorimeters, and so that these muon energies do not contribute to the reconstructed energies of hadronic jets. Next, electrons are reconstructed by matching ECAL deposits with electron tracks. Since most of the electron energy is lost within the ECAL volume, electrons are reconstructed by combining ECAL and tracker variables, with the track and ECAL clusters removed from the respective collections.

The remaining tracks and track-compatible ECAL and HCAL deposits are used to reconstruct charged hadrons. Neither neutral hadrons nor photons leave tracks in the tracker, but only the former leave deposits in the HCAL, allowing for reconstruction of and distinction between these particles.

5.1.2 Jet Recombination and Identification

Gluons and quarks cannot exist as free particles in the SM, or indeed in the NMSSM, and so they hadronise, causing a shower of hadrons in approximately the same direction. The resultant spray of hadrons is known as an jet, and may be reconstructed using the collinear- and infrared-safe anti-κ_{T} algorithm [86].

The anti-κ_{T} sequentially clusters input objects (protojets) into jets by starting with a high-p_{T} calorimeter seed deposit, protojet i, and merging it with the nearest deposit, protojet j, ascertained using the distance variable d_{ij} in Equation (5.1) defined in $(\eta - \phi)$ space. Similarly the distance between protojet i and the beam, d_{iB}, is given in Equation (5.2):

$$d_{ij} = min\left(p_{\mathrm{T}i}^{-2}, p_{\mathrm{T}j}^{-2}\right)\frac{\Delta R_{ij}^2}{R^2} \text{ , where } \Delta R_{ij}^2 = \left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2, \tag{5.1}$$

$$d_{iB} = p_{\mathrm{T}i}^2, \tag{5.2}$$

where $p_{\mathrm{T}i}$ and $p_{\mathrm{T}j}$ are the momenta of protojets i and j, with ΔR_{ij} representing the angular separation between the two protojets. R represents the radius parameter of the recombination algorithm. Standard jets as used in many CMS analyses such as those in this thesis are known as AK4 jets, meaning jets reconstructed using the anti-κ_{T} with radius parameter $R = 0.4$. Additionally larger-radius AK8 jets, with radius parameter $R = 0.8$, are sometimes used; notably as is the case in the experimental analysis in [2], considered in Chapter 7 of this thesis.

In the case where $d_{iB} < d_{ij}$ the seed particle i is merged with the beam. Otherwise protojets i and j are merged to form a new *pseudojet*. Once a pseudojet i becomes

sufficiently removed from any other pseudojet j such that $d_{iB} < d_{ij}$, it is considered a jet and the algorithm terminates.

In order to remove contributions to the reconstructed jet energy originating from PU events, a method is employed known as PileUp Per Particle Identification (PUPPI) [87]. Via the PUPPI method, the four-momenta of PF candidates entering the jet recombination algorithm are rescaled, with scale factor weights ranging from zero to one, depending on whether a particle originates from PU events or from hard scattering. This method proves useful in the case where jet substructure is examined, as it removes many of the neutral PU particles within the jets.

Additional jet energy corrections (JECs) are applied to simulated jets in order to align with the correct energy scale. This is driven by QCD multi-jet MC simulated events, with JECs having dependence on jet $|\eta|$ and p_{T} values. Moreover smearing is applied to the reconstructed jet energy for simulated events, to account for the worse resolution in data compared to simulation. Finally, jets must pass identification criteria, in order to reject any fake jets arising from mis-measurement and other effects.

5.1.3 Identification of Jets Containing Bottom Quarks

The analyses considered in this thesis involve the identification of bottom quark jets and jets containing a bottom quark-antiquark pair. In both cases identifying such jets relies on multivariate techniques which are described in this section.

Bottom Quark Jets

Bottom quarks may be identified in a detector due to their decay properties. Since the bottom quark has a relatively large decay lifetime compared with light flavour quarks, as the relevant CKM matrix elements are suppressed, its decay occurs a small distance away from the primary vertex (PV) in an event. As such, reconstructing displaced tracks allows for the identification of a secondary vertex, and thus a jet originating from a bottom quark; known as b-tagging.

The Combined Secondary Vertex v2 algorithm (CSVv2) [88] is used, combining the information from displaced track with that from reconstructed secondary vertices to generate a discriminator score. This score represents the likelihood that a given jet originated from the decay of a bottom quark.

The analysis in [15], considered in Chapter 6, places a threshold on the CSVv2 discriminator score such that the b-tagging efficiency is around 70%, with mis-tag rates for gluons and light flavour quarks being around $\sim 1\%$, derived from measurements of muon-jet events, $t\bar{t}$ events and QCD multijet events using data from CMS taken during Run II of the LHC [89].

Double-*b* Jets

In the case where an object, such as a Higgs boson, decaying to a bottom quark-antiquark pair is highly boosted, the angular separation ΔR between the two quarks is suppressed. Accordingly, as ΔR approaches 0.4, it becomes increasingly unlikely that the two quarks will be resolved as independent AK4 jets. Furthermore in the case where $\Delta R < 0.4$ these quarks cannot possibly be resolved into two such jets.

In this case it can be beneficial to consider an AK8 jet, formed with a larger radius parameter of 0.8, utilising a different algorithm designed for the identification of a bottom quark-antiquark pair. The two-prong substructure of this quark pair within the AK8 jet allows for separation of such jets from background objects.

In use since the beginning of Run-II of the LHC, a CMS-developed double-b-tagging algorithm [90] is employed for detecting AK8 jets containing bottom quark-antiquark pairs. This algorithm is based around the axis of each of the sub-jets within an AK8 jet. These variables along with others from the displaced tracks and secondary vertices form the 27 input variables, such as the number of secondary vertices within an AK8 jet and the distance between subjet axes, to a boosted decision tree (BDT). This multivariate approach generates a discriminator score which, akin to that of the CSVv2 algorithm, takes values in the range $[-1, 1]$.

Notably the double-b-tagging algorithm does not consider any jet mass variables, so as not to incur any inherent bias towards any particular mother particle mass. The double-b-tag discriminator score has very little correlation with jet p_{T} or the ΔR between bottom quarks, with the associated scale factor uncertainties applied to the weight of each event varying by only a few percent throughout the respective ranges of these variables [90].

5.2 Delphes Object Reconstruction

As discussed in Section 4.3, Delphes takes a parameterised approach to simulating the detector response and the reconstruction of physics objects. In this section we consider how Delphes reconstructs the particular objects of interest for the phenomenological work in Chapter 6.

5.2.1 Particle Flow Reconstruction

When reconstructing particles in an event Delphes takes a simplified approach [81] towards the PF algorithm [85]. Whilst in real experiments the momentum resolution of the tracker exceeds that of the calorimeters only up to a certain threshold, Delphes assumes this is always the case, and estimates the momenta of charged particles using the tracker information.

Two sets of PF 4-vectors are produced, representing tracks and calorimeter towers respectively, which are then used in reconstructing jets and jet-based variables. Each calorimeter tower is assigned E_{ECAL}, E_{HCAL}, $E_{\mathrm{ECAL,\,track}}$ and $E_{\mathrm{HCAL,\,track}}$. The former two represent the energy deposits in ECAL and HCAL respectively, with the latter being the energy deposits in the respective calorimeters originating from charged particles whose track has been reconstructed.

In order to then form PF tracks, the following variables are defined:

$$\begin{aligned}
\Delta_{\text{ECAL}} &:= E_{\text{ECAL}} - E_{\text{ECAL, track}}, \\
\Delta_{\text{HCAL}} &:= E_{\text{HCAL}} - E_{\text{HCAL, track}}, \\
E_{\text{Tower}}^{\text{eflow}} &:= \max(0, \Delta_{\text{ECAL}}) + \max(0, \Delta_{\text{HCAL}}).
\end{aligned} \tag{5.3}$$

Any reconstructed track then results in a PF track, with a PF tower created with energy $E_{\text{Tower}}^{\text{eflow}}$ if $E_{\text{Tower}}^{\text{eflow}} > 0$.

Via this method, particle flow tracks give a good resolution estimation of charged particles in an event. The particle flow towers on the other hand represent a combination of neutral particles, charged particles without a reconstructed track assigned and other contributions originating from the smearing of the calorimeter resolution; leading to a lower resolution for the towers. Furthermore this PF approach allows for high resolution when calculating jets and various energy sums [81], which will prove useful throughout this thesis.

Muons

Muons are reconstructed with some probability which vanishes outside the tracker acceptance region and below some momentum threshold. The reconstructed muon momentum is calculated via Gaussian smearing of the initial four-momentum vector, with the resolution parameterised as a function of p_{T} and η; all of these thresholds and parameters specified in the user's configuration parameters. Electrons, on the other hand, use a parameterised form of the combined tracker and ECAL reconstruction efficiency, with energy resolution being a combination of ECAL and tracker resolution [81]. Like muons however, electron reconstruction efficiency vanishes outside the tracker acceptance region and below some user-defined momentum threshold.

Photons

Photons are reconstructed solely using the ECAL information, neglecting conversion of photons into electron-positron pairs, with energy obtained using Equation (4.2) [81].

Hadronic Jets

Most importantly, given the all-hadronic final states in the analyses presented as part of this thesis, we consider the jet reconstruction within Delphes. Delphes produces jets in three different categories: *generated* jets, *calorimeter* jets and *PF* jets [81]. Generated jets are the generator-level jets; after showering and hadronisation has been performed, but with no simulated detector response or reconstruction applied. Calorimeter jets use the calorimeter resolution in Equation (4.2), whilst PF jets use the particle flow tracks and towers detailed in Section 5.2.1.

In the phenomenological work detailed in Chapter 6, PF jets are used, formed using the anti-k_T algorithm [86]. This parameter, along with the input variables for the reconstruction, isolation and identification of other objects, is set in the CMS Delphes configuration card, thus removing jets containing objects which have already been reconstructed as e.g. electrons and ensuring no objects are reconstructed twice.

***b*-tagging**

Finally, we have the tagging of bottom quark jets. Within Delphes b-tagging is purely parameterised: if a generated bottom quark or anti-quark is found within some angular distance ΔR of the axis of a jet, this jet is identified as a bottom quark with a probability dependent once again on the input parameters. Generally speaking, for the CMS b-tagging parameters, an isolated bottom quark within a standard jet usually has around a 70% chance of being correctly b-tagged. Furthermore, charm quarks and other flavours have approximately a 10% and 1% chance respectively of being incorrectly mis-identified as a bottom quark.

Owing to the high-p_{T} nature of the hadronic jets and the all-hadronic final state considered in the phenomenological work in Chapter 6 the reconstruction of other

objects is not considered. Furthermore whilst corrections for pile-up, jet energy scale and other effects are not considered, a validation exercise is performed, detailed in Section 6.4.7, which demonstrates that the event yields from the Delphes-based simulation are in agreement to within a few percent with the full simulation used in the experimental analysis being studied.

5.3 Common Variables: H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$

The analysis in [15] as well as the experimental analysis in Chapter 7 utilise two key jet-based variables, H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$. The calculations of these variables do not differ between the two analyses, however there are subtle differences in the jet reconstruction criteria between CMSSW and Delphes, which will be detailed later in this section.

H_{T} is defined as the scalar sum of the transverse momenta of the hadronic jets in an event which pass acceptance criteria such as p_{T}, η and various isolation requirements, as is detailed later in this section. Given the high-p_{T} jets expected in the SUSY decay cascades under investigation in this thesis, the mean H_{T} value is expected to be high in most regions of parameter space for signal MC events.

$H_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the norm of the two-dimensional vector sum of the transverse momenta of the jets in an event within the same acceptance regions as applied to H_{T}. Similar in construction, $E_{\mathrm{T}}^{\mathrm{miss}}$ includes all objects within these acceptance regions, such as leptons. Following the sole use of $H_{\mathrm{T}}^{\mathrm{miss}}$ throughout [15], this is used in lieu of $E_{\mathrm{T}}^{\mathrm{miss}}$ in all of the work presented in this thesis, noting that these quantities should be essentially equivalent in the case of an all-hadronic final state.

5.4 Phenomenology: Objects and Variables

In Chapter 6 we reinterpret the experimental analysis in Ref [15], which searches for simplified MSSM models, focusing on an all-hadronic final state with $E_{\mathrm{T}}^{\mathrm{miss}}$, utilising AK4 jet-based variables and applying a veto on events containing isolated

leptons and photons. This analysis additionally employs two variables, $\Delta\phi^*_{\text{min}}$ and α_T, defined in this section, designed to discriminate between signal processes featuring genuine $E_{\text{T}}^{\text{miss}}$ and QCD background with spurious $E_{\text{T}}^{\text{miss}}$ stemming from jet mis-measurement.

Crucially, whilst both of these $E_{\text{T}}^{\text{miss}}$-based variables are often employed to enhance signal process yields in comparison with background processes, QCD in particular, neither considers the overall *magnitude* of the $E_{\text{T}}^{\text{miss}}$ or $H_{\text{T}}^{\text{miss}}$. On this basis these variables should not require a large amount of $E_{\text{T}}^{\text{miss}}$, only that its source is at least one genuine invisible particle.

5.4.1 Hadronic Jets

Throughout the phenomenological work in Chapter 6 a standard "jet", prefaced either "AK4" or not at all, will refer to a hadronic jet constructed using the anti-k_T algorithm [86] with a radius parameter $R = 0.4$. Calculations of variables such as H_{T} and $H_{\text{T}}^{\text{miss}}$, defined later in this section, as well as the number of hadronic jets in an event all refer to jets of this type.

5.4.2 Leptons and Photons

A veto on events containing isolated leptons, photons and tracks is applied in Ref [15] in order to suppress large background contributions from processes decaying semi-leptonically, as well as unreconstructed leptons. It is important therefore, especially for data, that these objects have identification and isolation criteria applied. However, since the observed and background yields used were taken directly from Ref [15] and the signal mass points were simulated using Delphes rather than within CMSSW, the more subtle CMS-specific isolation and identification criteria were not applied to signal MC events.

5.4.3 $\Delta\phi^*_{\text{min}}$

$\Delta\phi^*$ is an angular variable which gauges to what extent a hadronic jet in an event responsible for observed $H_{\text{T}}^{\text{miss}}$ lies antiparallel to the vector sum of all other

hadronic jets in the plane perpendicular to the beam axis. The principle of this variable is that should this be the case then inaccurate measurement of the angle and energy of this jet is likely to be the source of the $H_{\mathrm{T}}^{\mathrm{miss}}$ measurement, rather than genuine $H_{\mathrm{T}}^{\mathrm{miss}}$ from an invisible, high-momentum particle escaping the detector.

Considering each hadronic jet in turn, calculating the respective $\Delta\phi^*$ and taking the minimum we obtain the $\Delta\phi^*_{\mathrm{min}}$ value for a given event. This value corresponds to the hadronic jet whose (mis-)measurement is most likely to lead to the observation of spurious $H_{\mathrm{T}}^{\mathrm{miss}}$, and so this value exceeding a given threshold indicates a degree to which the $H_{\mathrm{T}}^{\mathrm{miss}}$ might be declared genuine.

The calculation is detailed as follows, and is illustrated in Figure 5.1: First we define $\vec{p}_T^{\,i}$ as the transverse momentum vector of jet i, where $0 \leq i < N_{\mathrm{Jets}}$ and N_{Jets} is the number of hadronic jets. Now,

$$\vec{H}_{\mathrm{T}}^{\mathrm{miss},\,i} := \vec{H}_{\mathrm{T}}^{\mathrm{miss}} + \vec{p}_T^{\,i}. \tag{5.4}$$

Thus, we have

$$\Delta\phi^*_{\mathrm{min}} = \min(\Delta\phi(\vec{p}_T^{\,i}, \vec{H}_{\mathrm{T}}^{\mathrm{miss},\,i})), \tag{5.5}$$

where $\Delta\phi$ is the smallest azimuthal angle between the two vectors, given by

$$\Delta\phi(\phi_1, \phi_2) = |(\phi_1 - \phi_2) \bmod 2\pi| \tag{5.6}$$

Figure 5.1 shows a simplified example of an event featuring a tri-jet topology, with the beam axis pointing into the page. Here for each of the three jets the transverse momentum is indicated, along with the $H_{\mathrm{T}}^{\mathrm{miss}}$ calculated without the respective jet and the angular separation between the two. In this illustrative example we see that jet 1 has the smallest $\Delta\phi$, and thus we would have $\Delta\phi^*_{\mathrm{min}} = \Delta\phi_1$.

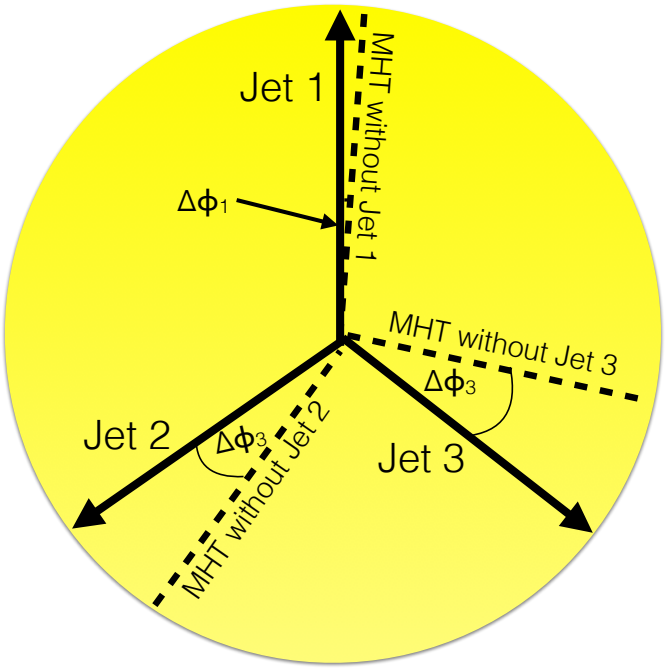


Figure 5.1: Example topology featuring three hadronic jets. The value for $\Delta\phi^*$ for each jet is calculated as the angular separation between the jet transverse momentum and the $H_{\mathrm{T}}^{\mathrm{miss}}$ calculated without the respective jet.

5.4.4 α_T

Unlike $\Delta\phi^*_{\mathrm{min}}$, α_T is a dimensionless variable, and its calculation includes the magnitude of the jet momentum, rather than azimuthal angle alone. It is, however, similar in purpose in that its intent is to distinguish between events with genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ and those with $E_{\mathrm{T}}^{\mathrm{miss}}$ observations occurring due to jet mis-measurement.

As defined in [45], in the simplest case where an event contains just two hadronic jets, α_{T} is calculated as follows:

$$\alpha_T = \frac{E_T^{j_2}}{M_T},$$

where $E_T^{j_2}$ is the transverse energy of the second-leading hadronic jet, sorted by transverse momentum, and

$$M_T = \sqrt{\left(\sum_{i=1}^{2} E_T^{j_i}\right)^2 - \left(\sum_{i=1}^{2} p_x^{j_i}\right)^2 - \left(\sum_{i=1}^{2} p_y^{j_i}\right)^2}.$$

For events with more than two hadronic jets, jets are combined to create *pseudo-jets* in such a way that the difference in p_{T} between these two pseudo-jets is minimised [46].

In the case of a final state consisting of two hadronic jets and no other particles, assuming a perfect detector with impeccable resolution, the jets would have equal and opposite momenta and so $\alpha_T = 0.5$. Should the jets be aligned back-to-back, but this time at least one jet is mis-measured, α_T would be expected to be less than 0.5.

On the other hand, in a scenario where two jets are produced along with some energetic, invisible particle, the jets would no longer be back-to-back and would recoil. Here α_T could exceed 0.5, allowing for this variable to distinguish between genuine and spurious $E_{\mathrm{T}}^{\mathrm{miss}}$.

5.5 Experimental Analysis: Objects and Variables

The experimental analysis detailed in chapter 7 studies some of the same variables and objects as the phenomenological work in chapter 6. However in many cases there are subtle differences in how these are calculated, reconstructed and corrected, which are detailed in this section.

5.5.1 Hadronic Jets

For both AK4 and AK8 hadronic jets, properties such as p_T, η and ϕ are taken from the unmodified respective default jet branch in the miniAOD files produced by CMSSW. Therefore, in the 2016 run year AK4 hadronic jets have Charged Hadron Subtraction (CHS) pile-up subtraction [91], whereas for the 2017 dataset AK4 jets utilise PUPPI pile-up subtraction instead. In the analysis detailed in Chapter 7 event yields corresponding to observed and MC simulated events from the 2016 and

2017 run years are considered separately, meaning no issues arise due to combination of samples from different run-years which feature different pile-up subtraction methods.

Otherwise the hadronic jets studied in this analysis are essentially the same as those detailed in section 5.4.1, requiring $p_{\mathrm{T}} > 40$ GeV, but in this case we require $|\eta| < 3.0$, since this is the maximal η range over which AK8 jets are considered by the double-b-tag discriminator.

5.5.2 Jet Identification Criteria

Since the calculations of H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$ consider central AK4 jets with p_T as low as 40 GeV, there is potential for this to include a small number of spurious jets reconstructed in error; one possible example being large energy deposits in HCAL originating from punch-through, caused by a very highly energetic lepton showering in, but not constrained within, ECAL.

Therefore, in order to reduce the likelihood of such contributions to H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$ calculations, a loose jet identification criteria is implemented, with any hadronic jets failing these criteria discarded. The loose jet ID requirements are defined in Table 5.1 [92]:

5.5.3 Jet Soft-drop Mass

In order to remove contributions to the reconstructed mass of a jet stemming from soft radiation, jet grooming techniques are often employed. In the analysis in Chapter 7 of this thesis the soft-drop algorithm is used [93], with the soft-drop mass referring to the jet mass after this grooming method has been performed.

Here we summarise the soft-drop grooming procedure. Starting with an AK8 jet j, the jet constituents are recombined using the Cambridge/Aachen recombination algorithm [94], which clusters pairs of particles with the smallest angular separation sequentially. The last stage of this clustering process is reversed, which leaves two

Loose Jet ID Criteria:	
For $\|\eta\| \leq 2.4$	
Attribute	Requirement
Neutral Hadron Fraction	< 0.99
Neutral Electromagnetic Fraction	< 0.99
Number of Constituents	> 1
Charged Hadron Fraction	< 0.99
Charged Multiplicity	> 0
Charged Electromagnetic Fraction	< 0.99
For $2.4 < \|\eta\| \leq 2.7$	
Attribute	Requirement
Neutral Hadron Fraction	< 0.99
Neutral Electromagnetic Fraction	< 0.99
Number of Constituents	> 1
For $2.7 < \|\eta\| \leq 3.0$	
Attribute	Requirement
Neutral Hadron Fraction	< 0.98
Neutral Electromagnetic Fraction	> 0.01
Number of Neutral Particles	> 2

Table 5.1: Detailed loose ID criteria as applied to hadronic AK4 jets considered in this analysis.

sub-jets, j_1 and j_2. We then define j as the sub-jet with higher p_{T} if these two sub-jets pass the following condition [93]:

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > \frac{1}{10} \cdot \frac{\Delta R_{1,2}}{R}, \tag{5.7}$$

where $R = 0.8$. This corresponds to the removal of a sub-jet if it contains less than 10 % of the total jet p_{T}, which has the effect of removing soft radiation. In the case where j is comprises a single particle, it is selected as the final jet. Otherwise, we undo the previous jet clustering stage and evaluate Equation (5.7) again.

This algorithm proves useful in the removal of mass contributions from soft radiation in the case of background jets. However as is seen in Section 7.2 its high power leads to a tendency to incorrectly remove hard PF candidates in jets originating from Higgs boson decays in signal events. This causes the final soft-drop mass to be very slightly lower than that of the decaying Higgs boson. However, we choose

to retain this grooming method when calculating the masses of AK8 jets in the analysis detailed in Chapter 7 in order to reduce background contributions from QCD multijet processes where soft radiation contributions could otherwise lead to erroneously high jet masses.

5.5.4 Object Isolation

In order to trace the origins of a reconstructed particle one requires some measure of how well isolated a particle is with respect to other reconstructed objects within an event. This allows for ascertaining whether for example a lepton is a prompt lepton, i.e. originating from the primary pp collision, a lepton stemming from the decay of an unstable particle (non-prompt), or a lepton formed via the hadronisation of a jet. Since the analyses presented in this thesis focus on an all-hadronic final state, lepton isolation is used in order to distinguish between these cases, allowing the removal of contributions from background processes involving leptons.

The typical isolation method applied to a reconstructed particle in many CMS analyses is PF relative isolation; calculated by summing the p_{T} of all charged hadrons (CH), neutral hadrons (NH) and photons (γ) within a cone of radius R about that particle and dividing by its p_{T}.

Only charged hadrons produced at the primary event vertex are considered, thus excluding particles produced from a different bunch crossing, in order to reject charged particles from pileup. However since neutral pileup particles cannot be associated with the primary vertex an estimated correction is applied, based upon an effective area (EA) [95], which will be discussed in Section 5.5.5.

The isolation method used in this analysis for identifying leptons is PF mini isolation [96], with standard PF relative isolation [96] used for photons and tracks. Whereas PF relative isolation uses a cone of fixed radius, in this case $R = 0.3$, about a reconstructed particle, mini isolation uses a variable cone radius, which decreases for high-p_{T} particles. This allows for increased efficiency in identify-

ing isolated particles in highly-boosted final states, where the angular separation between such particles and others in an event is suppressed.

We use this method to identify isolated leptons in order to remove events featuring leptons originating from particle decays, since the analyses in Chapters 6 and 7 focus on all-hadronic final states. Since these final states involve hadronic jets, which can contain non-isolated leptons, we only wish to remove events featuring isolated leptons.

The calculation of the PF isolation of an object l is defined as follows [96]:

$$I_{\mathrm{PF}}^{\mathrm{rel.}} := \frac{1}{p_{\mathrm{T}}^{l}}\Bigg(\sum_{\mathrm{PF\ (PV)}} p_{\mathrm{T}}^{\mathrm{CH}} \quad + \quad \sum_{\mathrm{PF}} \Big(p_{\mathrm{T}}^{\mathrm{NH}} + p_{\mathrm{T}}^{\gamma} \Big) \quad - \quad \frac{1}{2} \sum_{\mathrm{PU}} p_{\mathrm{T}}^{\mathrm{Neutral}} \Bigg), \tag{5.8}$$

considering only contributions within a cone of fixed or variable radius about the respective object, depending on whether relative or mini isolation is used. Here, $\sum_{\mathrm{PF}}$ denotes a sum over all particle flow candidates, $\sum_{\mathrm{PF\ (PV)}}$ a sum over particle flow candidates which stem from the primary vertex in an event and $\sum_{\mathrm{PU}}$ indicates a sum over all pileup contributions.

Thus the PF isolation $I_{\mathrm{PF}}^{\mathrm{rel.}}$ is defined as the fraction of energy contained within a cone around a lepton, relative to the energy of the lepton. Therefore for an isolated lepton, we have small values of $I_{\mathrm{PF}}^{\mathrm{rel.}}$, since the p_{T} sum of the PF candidates surrounding the lepton is lower than the lepton p_{T}. However for a lepton produced from a particle decay within a hadronic jet, $I_{\mathrm{PF}}^{\mathrm{rel.}}$ would be expected to be large, since the lepton p_{T} would be smaller than the total p_{T} sum of the remaining PF candidates from the jet.

We may also write the charged hadron, neutral hadron, photon and pileup isolation values individually, with each corresponding to the relevant sum in equation 5.8 above, as follows:

$$I_{\mathrm{PF}}^{\mathrm{rel.}} = I_{\mathrm{PF}}^{\mathrm{CH}} + I_{\mathrm{PF}}^{\mathrm{NH}} + I_{\mathrm{PF}}^{\gamma} - \frac{1}{2} I^{\mathrm{PU}} \tag{5.9}$$

In the case of lepton mini isolation the cone radius is p_{T}^{l}-dependent, and is defined as follows [96]:

$$\Delta R_{\mathrm{Max}} = \begin{cases} 0.2 & \text{if } p_{\mathrm{T}}^{l} < 50 \text{ GeV}, \\ 10\mathrm{GeV}/p_{\mathrm{T}}^{l} & \text{if } 50 \leq p_{\mathrm{T}}^{l} < 50 \text{ GeV}, \\ 0.005 & \text{if } p_{\mathrm{T}}^{l} \geq 200 \text{ GeV}. \end{cases} \tag{5.10}$$

5.5.5 Energy Density × Effective Area Corrections

In order to correct for pileup contributions within a cone around a reconstructed object we employ Effective Area (EA) corrections of the form $\rho \times$ EA, where ρ represents the energy density per unit area. The value of ρ is calculated in CMSSW per event as the median of the jet p_{T} per unit area distribution [95].

The EA values are η-dependent and represent an estimation of the area of the detector taken up by the lepton in the detector. These values are provided by CMS, and are derived using MC simulated events containing photons and hadronic jets, averaging over a range of pileup values [95].

These corrections apply to each of the charged hadron, neutral hadron, photon and pileup isolation values individually. For each isolation type the corrected value is defined as the maximum of zero and the uncorrected isolation with ρEA subtracted, shown in Equation (5.11).

$$\begin{aligned} I_{\mathrm{corr.}}^{\mathrm{CH}} &= \max(\, 0,\; I_{\mathrm{PF}}^{\mathrm{CH}} - \rho \cdot \mathrm{EA}_{\mathrm{CH}} \,) \\ I_{\mathrm{corr.}}^{\mathrm{NH}} &= \max(\, 0,\; I_{\mathrm{PF}}^{\mathrm{NH}} - \rho \cdot \mathrm{EA}_{\mathrm{NH}} \,) \\ I_{\mathrm{corr.}}^{\gamma} &= \max(\, 0,\; I_{\mathrm{PF}}^{\gamma} - \rho \cdot \mathrm{EA}_{\gamma} \,). \end{aligned} \tag{5.11}$$

5.5.6 Isolated Muons (Veto)

In this analysis isolated muons are considered only if they have $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.5$ and pass loose muon identification criteria. This loose muon ID is stored as a boolean flag in the miniAOD files, and is the equivalent of requiring that a muon be reconstructed as a PF muon and either a global muon or a tracker muon [96].

The isolation requirement for muons is that mini isolation $I_{\mathrm{PF}}^{\mathrm{mini}} < 0.2$, where $I_{\mathrm{PF}}^{\mathrm{mini}}$ is defined in Equation (5.8) featuring variable cone radius as defined in Equation (5.10).

5.5.7 Isolated Electrons (Veto)

The criteria electrons must pass in order to be considered are essentially the same as for muons. The p_{T} and $|\eta|$ requirements are identical, and electrons must also pass loose ID criteria. The ID requirements for electrons differ from those for muons, however, and are defined in Table 5.2.

In Table 5.2, the quantity $\sigma_{i\eta i\eta}$ refers to the energy-weighted variance in η of the ECAL crystals surrounding the cluster of 5×5 ECAL crystals, known as a supercluster, nearest to the electron track. The quantities $|\Delta\eta_{\mathrm{In}}|$ and $|\Delta\phi_{\mathrm{In}}|$ represent the differences in η and ϕ respectively between the electron track and the ECAL supercluster [95].

Electron isolation also uses essentially the same mini isolation algorithm as for muons, with a tighter threshold: for electrons to be considered isolated, they must have $I_{\mathrm{PF}}^{\mathrm{mini}} < 0.1$. A subtle difference is that the charged hadron, neutral hadron and photon contributions to $I_{\mathrm{PF}}^{\mathrm{mini}}$ incur $\rho \times$ EA corrections, in order to remove effects of pileup.

5.5.8 Isolated Photons and Tracks(Veto)

Isolated photons and tracks are also considered in the analysis in Chapter 7, with both required to pass similar isolation criteria. The isolation I_{PF} of both is cal-

Loose Electron ID Criteria:	
For Barrel Region ($\|\eta_{\mathrm{SC}}\| < 1.479$)	
Attribute	Requirement
$5 \times 5\ \sigma_{i\eta i\eta}$	< 0.011
$\|\Delta\eta_{\mathrm{In}}\|$	< 0.00477
$\|\Delta\phi_{\mathrm{In}}\|$	< 0.222
H/E	< 0.298
d0 [cm]	< 0.05
dZ [cm]	< 0.10
$\|(\frac{1}{E_{\mathrm{ECAL}}} - \frac{1}{p_{\mathrm{track}}})\|$	< 0.241
Missing Hits (Inner Tracker)	≤ 1
Passes Conversion Veto	Yes
For Endcap Region ($1.479 \leq \|\eta_{\mathrm{SC}}\| < 2.5$)	
Attribute	Requirement
$5 \times 5\ \sigma_{i\eta i\eta}$	< 0.0314
$\|\Delta\eta_{\mathrm{In}}\|$	< 0.00868
$\|\Delta\phi_{\mathrm{In}}\|$	< 0.213
H/E	< 0.101
d0 [cm]	< 0.10
dZ [cm]	< 0.20
$\|(\frac{1}{E_{\mathrm{ECAL}}} - \frac{1}{p_{\mathrm{track}}})\|$	< 0.14
Missing Hits (Inner Tracker)	≤ 1
Passes Conversion Veto	Yes

Table 5.2: Detailed loose ID criteria as applied to electrons considered in this analysis.

culated using the PF isolation algorithm; identical to the mini relative isolation algorithm, but with a fixed radius parameter of, in this case, $\Delta R < 0.3$. Both photons and tracks are considered isolated if $I_{\mathrm{PF}} < 0.1$. Events are vetoed if they contain an isolated photon with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$, or an isolated track with $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$.

5.6 Summary

In this chapter we have examined the measurement variables, as well as the definitions and acceptance criteria for reconstructed particles, which will be used throughout this thesis. Some of these variables, such as H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$, are used by both of the experimental analyses in Chapters 6 and 7, albeit with minor differences in definition. Certain quantities on the other hand, as will be seen in the following chapters, such as $\Delta\phi^{*}_{\mathrm{min}}$ and AK8 jet soft-drop mass, are specific to a particular analysis.

6 | Phenomenological Interpretation of Jets+$E_{\mathrm{T}}^{\mathrm{miss}}$ Analysis

The low-$E_{\mathrm{T}}^{\mathrm{miss}}$ NMSSM scenarios under consideration in this thesis typically have weaker constraints imposed by experimental searches targeting simplified MSSM models which produce larger amounts of $E_{\mathrm{T}}^{\mathrm{miss}}$ [37]. However, the searches considered in [37] were performed mainly on data from Run I of the LHC, with lower bounds placed on sparticle masses in the $1 - 1.5$ TeV range. In order to gain understanding of how these models may be tested using current search efforts in Run II, the CMS α_{T}-based general-purpose BSM analysis in Ref [15] was re-examined in the context of the NMSSM scenarios under investigation in this thesis.

This analysis was chosen since it focuses on an all-hadronic final state, with the event yields split into bins depending on the number of hadronic jets, b-tagged jets, total H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$, given the NMSSM decay cascades considered in this thesis are expected to produce large numbers of hadronic jets, and include two Higgs bosons each decaying to a bottom quark-antiquark pair. Results of this reinterpretation are based upon data collected from pp collisions during the 2016 run year at the CMS experiment, at an integrated luminosity of 35.9fb^{-1}.

Firstly however, we define a number of mass points, starting from the eight benchmark points in Table 2.2 and varying the squark, gluino, NLSP and LSP masses in order to explore how the expected yields from each of these mass points compare with those of SM background processes and observed yields from data. This then allows for a measure of experimental sensitivity to be defined, such as the ratio of the expected signal yield, s, to the overall expected event yield, denoted $s/\sqrt{s+b}$, where b represents the expected SM background yield.

Mass Scan	$M_{\tilde{q}}$ [GeV]	$M_{\tilde{g}}$ [GeV]	$M_{\tilde{\chi}_1^0}$ [GeV]	$M_{\tilde{\chi}_2^0}$ [GeV]	$M_{\tilde{t},\tilde{b}}$ [GeV]
BM1	$1200 \rightarrow 3000$	$M_{\tilde{q}} + 10$	$\sim 1 \rightarrow \{M_{\tilde{q}} - 150\}$	$\sim M_{\tilde{\chi}_1^0} + 125$	decoupled
BM2	$1200 \rightarrow 3000$	$M_{\tilde{q}} - 200$	$\sim 1 \rightarrow \{M_{\tilde{g}} - 150\}$	$\sim M_{\tilde{\chi}_1^0} + 125$	decoupled
BM3	$1200 \rightarrow 3000$	$M_{\tilde{q}} + 10$	$\sim 1 \rightarrow \{M_{\tilde{t}} - 330\}$	$\sim M_{\tilde{\chi}_1^0} + 125$	$M_{\tilde{t}} = M_{\tilde{q}} - 250$
BM4	$1200 \rightarrow 3000$	$M_{\tilde{q}} + 10$	$\sim 1 \rightarrow \{M_{\tilde{b}} - 150\}$	$\sim M_{\tilde{\chi}_1^0} + 125$	$M_{\tilde{b}} = M_{\tilde{q}} - 250$
BM5	$1200 \rightarrow 3000$	$M_{\tilde{q}} - 200$	$\sim 1 \rightarrow \{M_{\tilde{t}} - 330\}$	$\sim M_{\tilde{\chi}_1^0} + 125$	$M_{\tilde{t}} = M_{\tilde{g}} - 250$
BM6	$1200 \rightarrow 3000$	$M_{\tilde{q}} - 200$	$\sim 1 \rightarrow \{M_{\tilde{b}} - 150\}$	$\sim M_{\tilde{\chi}_1^0} + 125$	$M_{\tilde{b}} = M_{\tilde{g}} - 250$

Table 6.1: Table showing various mass ranges across the six 2D mass scans. $M_{\tilde{q}}, M_{\tilde{g}}$ are varied together in steps of 100 GeV and $M_{\tilde{\chi}_1^0,\tilde{\chi}_2^0}$ in steps of 50 GeV

6.1 Discrete Two-Dimensional Mass Scans

We start from the BPs in Table 2.2, which were defined in [37] in order to demonstrate candidate mass hierarchies which would generate small $E_{\mathrm{T}}^{\mathrm{miss}}$ and which have evaded detection by SUSY searches using data from Run I of the LHC. Mass scans are developed, by varying respective mass values in discrete steps and generating MC simulated events at each step, which envelop and characterise these points in order to determine to what extent such models may escape detection from search efforts such as the analysis in [15]. In all cases MC simulated events are generated at parton-level, with the decays, hadronisation and detector response calculated for each point in these scans.

Considering each of the existing BPs in Table 2.2, where $M_{\tilde{\chi}_2^0} = 130$ and $M_{\tilde{\chi}_1^0} = 3$ GeV, a two-dimensional mass grid is constructed as follows. The mass gaps $M_{\tilde{q}} - M_{\tilde{g}}$ and $M_{\tilde{\chi}_2^0} - M_{\tilde{\chi}_1^0}$ are kept constant, with $M_{\tilde{q},\tilde{g}}$ and $M_{\tilde{\chi}_1^0,\tilde{\chi}_2^0}$ now treated as two independent parameters. These mass values are then scanned across, generating MC simulated events at each point, by increasing $M_{\tilde{q}}, M_{\tilde{g}}$ together in steps of 100 GeV and $M_{\tilde{\chi}_1^0,\tilde{\chi}_2^0}$ together in steps of 50 GeV, with the ranges detailed in Table 6.1.

However, it may be noted that BP1 and BP2 are essentially the same, but for $M_{\tilde{q},\tilde{g}}$ being 400 GeV heavier in the latter. Therefore, one would find that a mass scan about BP1 would encapsulate BP2 anyway. This is also the case for BP3 with respect to BP4, however, BP5 to BP8 transform into four independent scans,

since for BP5 and BP7 the stop squark is involved in the decay cascades but not the sbottom squark. For BP6 and BP8 the converse applies.

Whilst it is expected that the sensitivity to these NMSSM scenarios be lowest for the lightest LSP, due to the suppressed $H_{\mathrm{T}}^{\mathrm{miss}}$ distribution, the masses of the NLSP and LSP in each scan are increased as high as is possible whilst remaining below the mass of the lighter of the squark and gluino. This is done in order to cover the entirety of the available sparticle mass range, such that it is possible to compare the same NMSSM scenario with a range of LSP masses.

We therefore define six independent discrete mass scans: BM1 based around the BP1-BP2 benchmark points, BM2 around BP3-BP4, and BM3–BM6 based around each of the remaining benchmark points BP5–BP8 respectively.

For the first two scans the NLSP mass is increased up to just below the lighter of the squark and gluino masses whilst still allowing for on-shell decay. For the remaining four scans, where the respective stop or sbottom squark is non-decoupled, its mass is set to be 250 GeV lower than the lightest of the squark and gluino, such that the gluino may still decay into the relevant third generation squark, along with an appropriate quark. In these cases, the NLSP mass may still be increased, so long that the involved third generation squark may still decay in an on-shell fashion into its respectively flavoured quark and an NLSP.

6.2 Cross-Section Calculation: PROSPINO

Ordinarily, for processes generated at a sufficiently high order, the production cross-section may be taken from Pythia after jet matching has been performed. However, as mentioned in section 4.1.1 only LO simulation is possible for the mass points in the NMSSM scenarios considered in this thesis.

As such, in addition to the inclusion of up to two extra jets at parton level, the cross-section reported by Pythia is discarded in favour of a value calculated separately at NLO using PROSPINO, "a program for the PROduction of Supersymmetric Particles In Next-to-Leading Order" [97].

PROSPINO is a Fortran package designed for calculating the cross-sections of squark-squark, squark-antisquark and squark-gluino final states at NLO within various SUSY models [98], along with other final states not considered in this thesis.

By default, PROSPINO uses five-fold degeneracy when calculating the cross-sections of processes involving squarks. However the mass points considered in this thesis typically include the initial production of only the first four squark flavours, since the stop and sbottom squarks are either decoupled or are only present further down in the decay cascade and whose direct production cross-section is accordingly neglected. As such, and since the cross-section of such third generation squarks is far smaller than that of other squarks of similar mass [98], the PROSPINO configuration was modified to only consider the first two generations of squark in the final state.

6.2.1 Validation of PROSPINO Cross-Section Values

In order to check the validity of the NLO cross-section values acquired using PROSPINO, we compare with equivalent official figures taken from the LHC Cross-Section Working Group (LHCXSWG) [99], calculated at NNLO+NNLL. Since these LHCXSWG calculations consider the first five squark flavours to be degenerate in mass, the same conditions are applied in Prospino for the purpose of this comparison.

Furthermore the LHCXSWG cross-section values are calculated with all other particle masses decoupled, meaning for example that the squark-squark production cross-section will not consider the production of heavier gluinos decaying into squarks. Therefore for the purposes of estimating the production cross-sections for mass points considered in this thesis, the LHCXSWG values are not appropriate. In order to directly compare these with values generated using PROSPINO, the same conditions are applied.

	Order	$\tilde{q}\tilde{q}$ [fb]	$\tilde{q}\tilde{g}$ [fb]	$\tilde{g}\tilde{g}$ [fb]
Prospino	NLO	7.09×10^{-3}	2.26×10^{-1}	4.24×10^{-2}
LHCXSWG	NNLO+NNLL	5.98×10^{-3}	2.23×10^{-1}	4.39×10^{-2}

Table 6.2: Production cross-section values in fb for various final states with $M_{\tilde{q}} =$ 2600 GeV and $M_{\tilde{g}} = 2610$ GeV, compared between PROSPINO at NLO and the official LHCXSWG figures calculated at NNLO+NNLL.

Table 6.2 shows a comparison for three sub-processes considered between PROSPINO at NLO and the official LHCXSWG values calculated at NNLO+NNLL. The sub-processes considered are squark-antisquark, squark-gluino and gluino-gluino production, for an example mass point with $M_{\tilde{q}} = 2600$ GeVand $M_{\tilde{g}} = 2610$ GeV. This high squark mass is chosen in order to consider the worst case scenario, where any discrepancies are expected to hold the most significance.

Following the calculation methods for the LHCXSWG cross-section values, the following criteria are applied. For squark-antisquark production the gluino is forced to be decoupled. Similarly the gluino-gluino cross-section is calculated with all squark masses decoupled. In all cases, the stop squark and all other sparticles are decoupled.

Here we see in general good agreement between PROSPINO and the official LHCXSWG figures for the squark-gluino and gluino-gluino production cross-section values, being compatible to within a few percent. The squark-antisquark production cross-section raises some discrepancy however, with the PROSPINO calculation giving a value approximately 20% higher than the official figure. However, this cross-section is around an order of magnitude smaller than those for squark-gluino and gluino-gluino production, and this discrepancy is well within the systematic uncertainty of around 40 % quoted on the LHCXSWG figure for this particular production cross-section value [99].

Therefore, considering the acceptable agreement between PROSPINO at NLO and the LHCXSWG cross-section values computed at NNLO+NNLL, we use PROSPINO to calculate the initial state production cross-section values at NLO for the mass

points considered in the analysis work in this thesis, and use the largest discrepancy to motivate an approximation of 20 % for the relevant systematic uncertainty on these cross-section values.

6.3 Kinematic Event Selection

The considered experimental analysis [15] contains many measurement bins for various observables in the data, background and signal channels; in particular for the number of hadronic jets (N_{jets}), the number of b-tagged hadronic jets ($N_{b\text{-jets}}$), H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$, each of which is defined in Chapter 5. However this analysis contains 254 measurement bins, many of which require very few hadronic jets and low H_{T} and are thus not relevant when searching for the high-H_{T}, low-$E_{\mathrm{T}}^{\mathrm{miss}}$ NMSSM scenarios considered for this thesis, typically featuring high-H_{T}, and so a subset is chosen.

Given the SUSY decay cascades in these NMSSM scenarios, shown in Figure 2.6, end in two LSPs and two Higgs bosons, each decaying to a bottom quark-antiquark pair, we focus on a $b\bar{b}b\bar{b}$+$E_{\mathrm{T}}^{\mathrm{miss}}$ final state, with plenty of jets from both cascades. We consider only the highest N_{jets} bin, the one for which $N_{\mathrm{jets}} \geq 6$, due to the large number of hadronic jets produced in the decay cascades in these NMSSM scenarios, anticipating four bottom quark jets from the two Higgs bosons, as well as at least two hadronic jets from the squark and gluino decays. Since if the Higgs bosons decaying to bottom quark pairs are sufficiently boosted the angular separation between the quarks in each pair will become small enough that it is no longer possible to resolve them independently, we consider bins corresponding to the cases where $N_{b\text{-jets}} = 2$, $N_{b\text{-jets}} = 3$ and $N_{b\text{-jets}} \geq 4$.

We are primarily interested in topologies featuring high jet multiplicity. Whilst it is possible for the decay cascades in the NMSSM scenarios under investigation to feature a lower number of jets, for example if neither Higgs boson decays to a bottom quark-antiquark pair, we consider only the case where both Higgs bosons decay via this channel. Since for squark and gluino masses greater than 1 TeV such

topologies would generate high H_{T}, we focus on the uppermost $H_{\mathrm{T}} \geq 1200$ GeV bin, with the exception of the case where we have four or more b-tagged hadronic jets, where the only H_{T} bin is an inclusive one, requiring $H_{\mathrm{T}} \geq 400$ GeV, due to low event statistics.

A cut on $\Delta\phi^{*}_{\mathrm{min}}$ is applied to reduce the QCD background, since this is designed to distinguish between events featuring a genuine source of $E_{\mathrm{T}}^{\mathrm{miss}}$ and those featuring spurious $E_{\mathrm{T}}^{\mathrm{miss}}$. Additional cuts as in Ref [15] regarding vetoing events containing isolated leptons and photons are also performed, in order to reduce background contributions from leptonically decaying processes, as well as removing events which contain hadronic jets aligned closely with the beam axis in order to maintain high $H_{\mathrm{T}}^{\mathrm{miss}}$ resolution, since such jets lie outside the tracker acceptance region. The event selection is therefore detailed as follows:

- At least 6 hadronic jets, where any jet must have $p_{\mathrm{T}} \geq$ 40 GeV in order to ensure $\geq 95\%$ jet reconstruction efficiency [15].
- $N_{b\text{-jets}} = 2$, $N_{b\text{-jets}} = 3$, $N_{b\text{-jets}} \geq 4$, i.e., separate bins.
- $N_{b\text{-jets}}$, H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$ binning defined in table 6.4.
- $H_{\mathrm{T}} \geq 1200$ GeV for events where $N_{b\text{-jets}} = 2, 3$, or $H_{\mathrm{T}} \geq 400$ GeV where $N_{b\text{-jets}} \geq 4$.
- $\Delta\phi^{\star} \geq 0.5$.

Event vetoes are defined such that events will fail should they contain any of the following isolated objects:

- photons with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$,
- electrons and muons with $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$,
- jets with $p_{\mathrm{T}} > 40$ GeV and $|\eta| > 2.4$,

H_{T} [GeV]	$200-250$	$250-300$	$300-350$	$350-400$	$400-900$	$900-\infty$
α_{T}	>0.65	>0.6	>0.55	>0.53	>0.52	>0

Table 6.3: Table detailing the H_{T}-dependent α_{T} cuts.

Bin	H_T [GeV]	H_T^{miss} [GeV]	Data Yield	Expected Background Yield
$=2$b	1200–∞	200–400	0	2.51 ± 1.02
$=2$b	1200–∞	400–600	0	1.65 ± 0.44
$=2$b	1200–∞	600–900	2	0.62 ± 0.32
$=2$b	1200–∞	900–∞	0	0.19 ± 0.18
$=3$b	1200–∞	200–400	1	0.40 ± 0.16
$=3$b	1200–∞	400–600	0	0.25 ± 0.08
$=3$b	1200–∞	600–900	1	0.09 ± 0.04
$=3$b	1200–∞	900–∞	0	0.02 ± 0.02
≥ 4b	400–∞	200–∞	4	2.46 ± 0.70

Table 6.4: Observed and MC background yields for each of the bins used in this analysis, taken from [15].

where the relevant isolation definitions are detailed fully in Chapter 5. These lepton vetoes are intended to reject SM background processes containing leptonically decaying objects which may give rise to $H_{\mathrm{T}}^{\mathrm{miss}}$, for example $W \rightarrow l\nu$, where l represents an electron or a muon.

Finally a cut on α_{T} is included, with H_{T}-dependent thresholds shown in Table 6.3. We note here that owing to the $H_{\mathrm{T}} > 1200$ GeVrequirement across the $N_{b\text{-jets}} = 2, 3$ bins, this α_{T} requirement applies only to the $N_{b\text{-jets}} \geq 4$ bin.

6.3.1 Signal, Background and Observed Event Yields

Table 6.4 contains the observed and MC background yields taken directly from [15] for each of the $N_{b\text{-jets}}$ and $H_{\mathrm{T}}^{\mathrm{miss}}$ bins, satisfying the kinematic event selection criteria.

The event yields shown in Table 6.4 demonstrate good agreement between the expected background yields and the observations in data, with no statistically significant excesses found given the small number of observed events and thus high relative statistical uncertainty. The observed yields and background estimations are then used to calculate lower bounds on the sparticle masses given the signal yields for each mass point in each mass scan, in order to explore the sensitivity of

this analysis to the low-$E_{\mathrm{T}}^{\mathrm{miss}}$ NMSSM scenarios considered in this thesis, and thus to what extent these scenarios may escape sensitivity of such a general purpose jets+$E_{\mathrm{T}}^{\mathrm{miss}}$ analysis.

6.4 Distributions of Key Variables for Signal and Background Processes

We examine the observable properties of these BPs, in order to explore which event selections have the greatest impact on the experimental sensitivity. Additionally we also consider the MC *truth* values, i.e., the generated values without detector simulation being applied, for quantities such as b jet angular separation, since this quantity taking a value below the resolution of the detector will have a large effect on the efficiency of bottom quark tagging and the ability to resolve both b jets stemming from the decay of each Higgs boson.

First consider the BM1-BM6-type mass scans, taking from each two mass points where one has the lightest 3 GeV LSP and the other a mid-range 953 GeV LSP representative of a simplified MSSM-like model, choosing an example squark mass of 2 TeV, close to the current experimental limits for simplified MSSM-like models [15].

In this section we see the distributions of quantities of interest for all events, before any event selection is applied, normalising to unity for comparison. We then see how the respective event selection criteria in this analysis affect the expected signal event yields in the low- and high-$M_{\tilde{\chi}_1^0}$ regions respectively. These distributions are then overlayed with the corresponding normalised distributions for background processes containing QCD multijet events and background contributions from top quark pair production.

Additionally, a simplified MSSM-like model is presented for comparison. This scenario is derived from the BM1-type scan, with the difference being that the NLSP is dropped, with the squarks and gluinos decaying instead directly to the LSP and

hadronic jets. Thus no Higgs bosons are produced and the LSP momentum is no longer suppressed, since the light flavour hadronic jets are lighter than the LSP, removing the possibility of a low $E_{\mathrm{T}}^{\mathrm{miss}}$ and Higgs boson enriched scenario. In this model the squark mass is set at 2 TeV and the effective LSP has mass 3 GeV.

6.4.1 Total scalar H_{T}

A dominant feature of many SUSY cascades is high H_{T}, due to the large number of jets produced in the decay cascades. As shown in Figure 6.1 the light-LSP scenarios featuring heavy squarks and gluinos deliver an H_{T} distribution with mean well over 2 TeV.

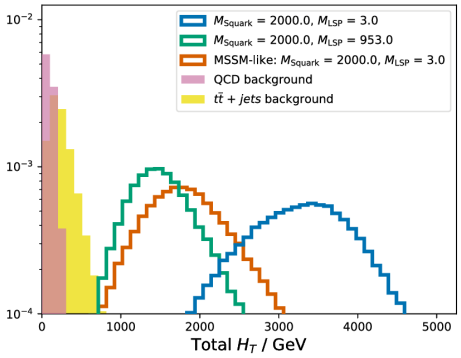
Figure 6.1 shows the H_{T} distributions for two example points from each of the six mass scans. Whilst it is clear that the mean H_{T} for both QCD multijet and top pair production background processes is far lower than that for the signal processes, their respective cross-sections are much higher. Therefore, whilst the peaks of these distributions are well separated, it would be expected for the tails of both background processes to still be significant compared with the signal processes when considering the event yields in an analysis.

In fact, considering the fraction of events passing the 1200 GeV minimum H_{T} requirement across the BM1-type mass scan range, shown in figure 6.2, it becomes clear that for many of the mass points essentially all events pass this cut.

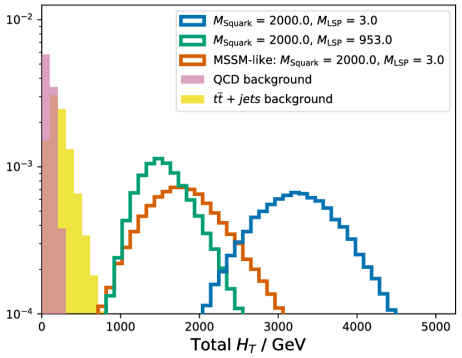
However, it is also apparent that the mean H_{T} is much lower where the LSP mass approaches that of the squarks and gluinos. As such, as seen for the regions near the diagonal in Figure 6.2, in this case the fraction of events passing the $H_{\mathrm{T}} > 1200$ GeV requirement is much lower. This is the case since the mass gaps in the decay cascade are reduced and so the p_{T} of any emitted quark is suppressed.

6.4.2 $H_{\mathrm{T}}^{\mathrm{miss}}$

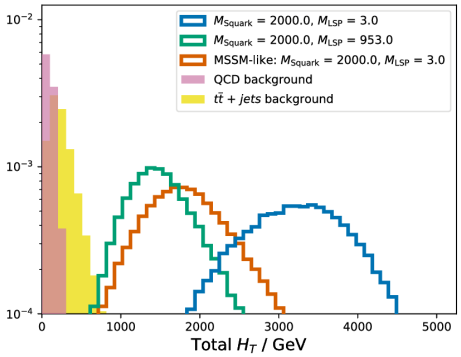
Of course the existence of these LSP with non-zero momenta dictates there must also be some missing net transverse energy, so long that the two LSPs do not have equal and opposite momenta.



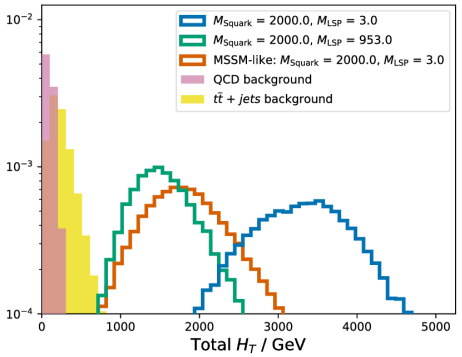
(a) BM1-type Mass Scan.



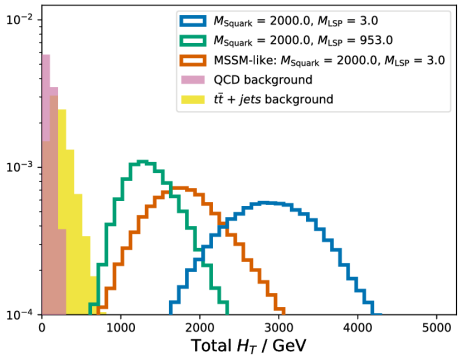
(b) BM2-type Mass Scan.



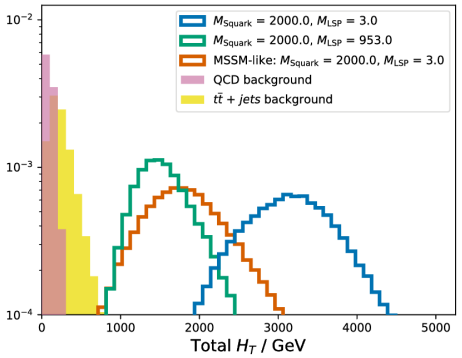
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.

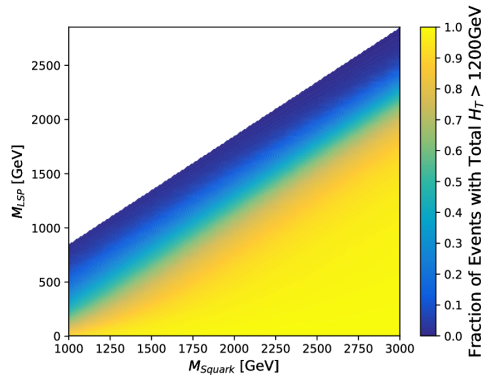


(e) BM5-type Mass Scan.

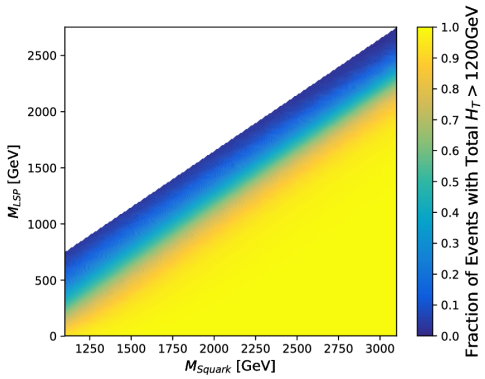


(f) BM6-type Mass Scan.

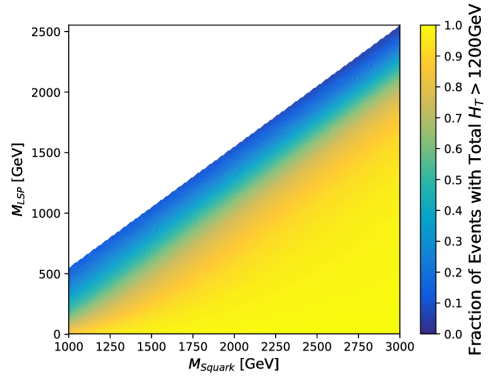
Figure 6.1: Normalised H_{T} distributions for $M_{\mathrm{LSP}} = 3$ GeV and $M_{\mathrm{LSP}} = 953$ GeV, where $M_{\tilde{q}} = 2$ TeV, in the BM1-BM6-type scan, compared with QCD and $t\bar{t}$ background processes and an MSSM-like scenario with a light 3 GeV LSP. We note that the minimum H_{T} threshold is 1200 GeV.



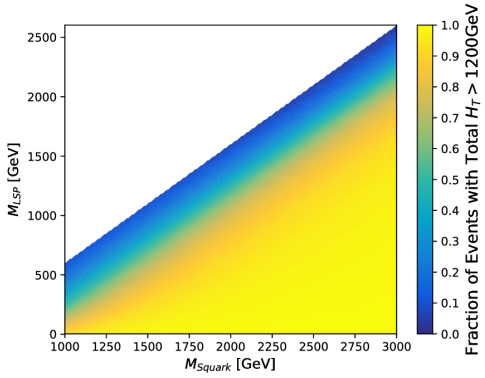
(a) BM1-type Mass Scan.



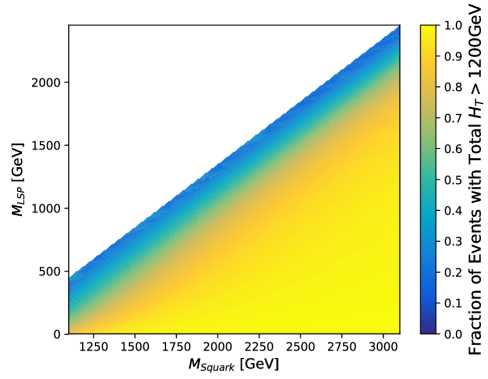
(b) BM2-type Mass Scan.



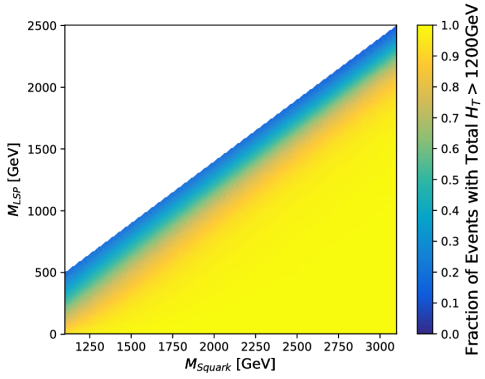
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.



(e) BM5-type Mass Scan.



(f) BM6-type Mass Scan.

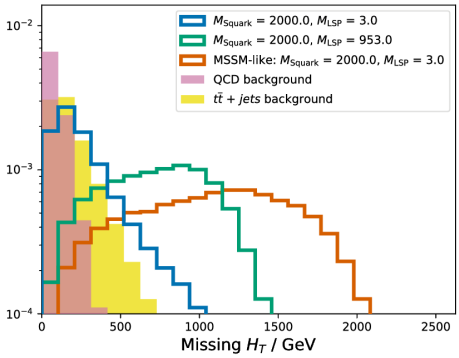
Figure 6.2: Fraction of events with total $H_{\mathrm{T}} > 1200\,\mathrm{GeV}$ for the BM1-BM6-type mass scans.

In Figure 6.3 rather distinct distributions between the two mass points may be observed. The heavy LSP scenarios give a fairly wide spread of $H_{\mathrm{T}}^{\mathrm{miss}}$ values due to the presence of such a heavy, boosted and invisible LSP. The light-LSP scenarios however suppress this, since the considerably heavier Higgs boson will inherit most of the momentum from the NLSP decay, leaving a soft LSP.

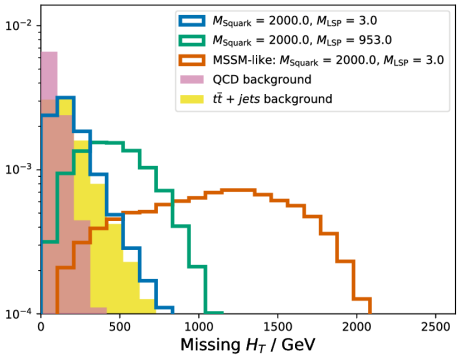
Even in these light-LSP scenarios, there are still events whose $H_{\mathrm{T}}^{\mathrm{miss}}$ is quite high. This larger $H_{\mathrm{T}}^{\mathrm{miss}}$ can arise via Higgs bosons decaying to final states other than a bottom quark-antiquark pair, though the branching fractions for decays such as $H \rightarrow ZZ^{*} \rightarrow \nu\bar{\nu}\nu\bar{\nu}$ are very small. A more likely phenomenon is for one or more of the bottom quarks stemming from the decays of the Higgs bosons to decay into a charm quark and an electron or muon, along with the appropriately-flavoured neutrino, with the neutrino momentum contributing to the overall $H_{\mathrm{T}}^{\mathrm{miss}}$. Additionally, hadronic Higgs boson decay final states such as $b\bar{b}$ are difficult to reconstruct owing to resolution and, given the higher number of jets compared with, say, a cleaner $H \rightarrow \gamma\gamma$ decay channel, mean more broad distributions in $H_{\mathrm{T}}^{\mathrm{miss}}$ and H_{T}. Since we are using the uppermost $H_{\mathrm{T}} > 1200$ GeV bin, with quite coarse $H_{\mathrm{T}}^{\mathrm{miss}}$ binning, any effect on H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$ resolution is expected to be minimal, and is not considered directly.

As may be seen most notably in the lower regions of Figures 6.3a and 6.3b, the $H_{\mathrm{T}}^{\mathrm{miss}}$ decreases drastically in the limit of a light-LSP in the NMSSM, with the peak well below 200 GeV, the minimum $H_{\mathrm{T}}^{\mathrm{miss}}$ requirement used in this analysis. It is clear for these areas of mass space many of the events are lost due to $H_{\mathrm{T}}^{\mathrm{miss}}$ cuts, thus decreasing experimental sensitivity to this type of model.

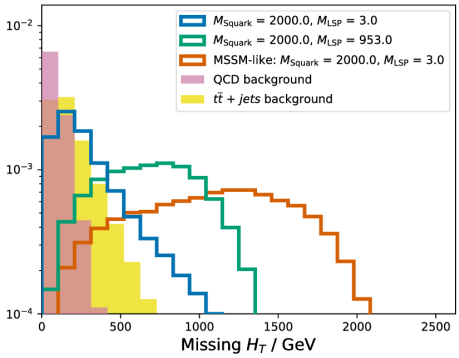
Additionally, as seen in the colour map plots in figure 6.4, the fraction of events with $H_{\mathrm{T}}^{\mathrm{miss}}$ greater than the lower edge of the minimum $H_{\mathrm{T}}^{\mathrm{miss}}$ bin, 200GeV, rises considerably as the LSP mass increases. As the LSP mass approaches that of the squarks and gluinos however, towards the upper edge of the coloured region, this fraction begins to drop since the mass gaps in the decay cascade shrink, meaning



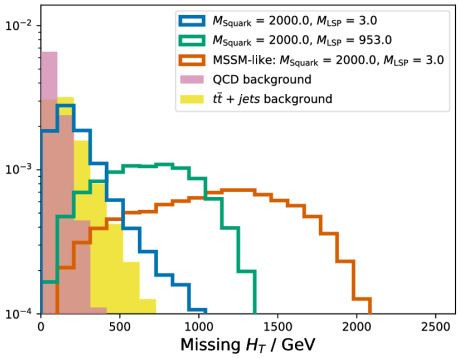
(a) BM1-type Mass Scan.



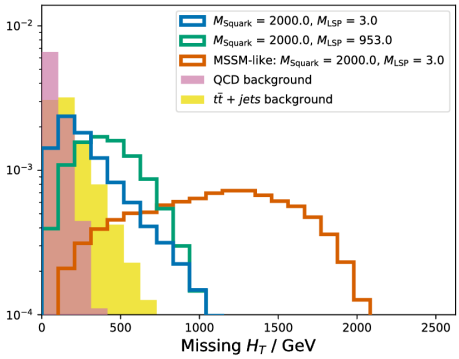
(b) BM2-type Mass Scan.



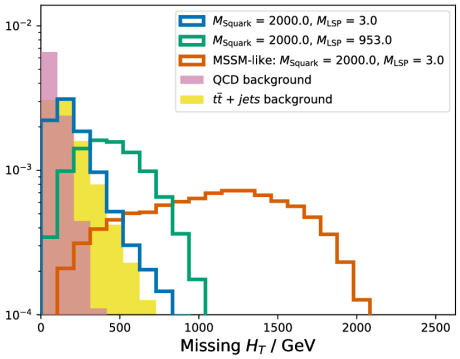
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.

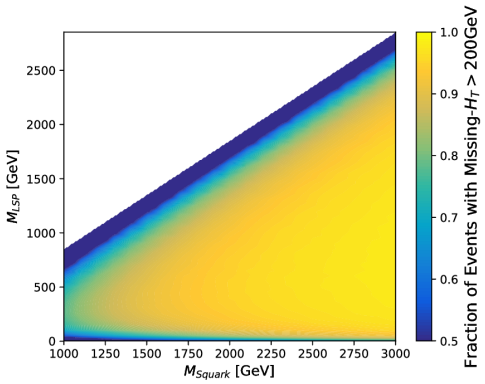


(e) BM5-type Mass Scan.

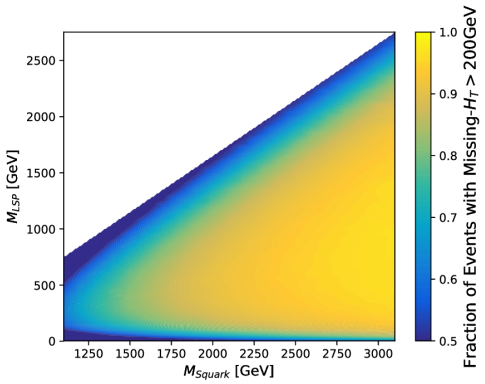


(f) BM6-type Mass Scan.

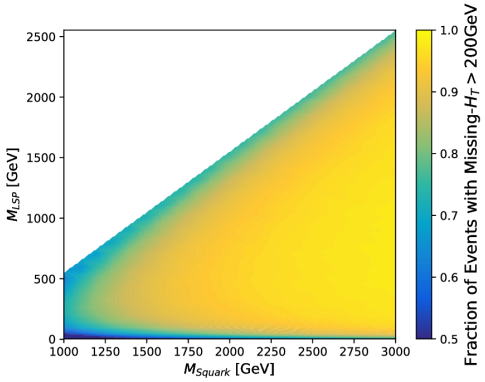
Figure 6.3: Normalised H_{T} distributions for $M_{\mathrm{LSP}} = 3$ GeV and $M_{\mathrm{LSP}} = 953$ GeV, where $M_{\tilde{q}} = 2$ TeV, in the BM1-type scan, compared with QCD and $t\bar{t}$ background processes and an MSSM-like scenario with a light 3 GeV LSP. We note that the minimum $H_{\mathrm{T}}^{\mathrm{miss}}$ threshold considered in this analysis is 200 GeV.



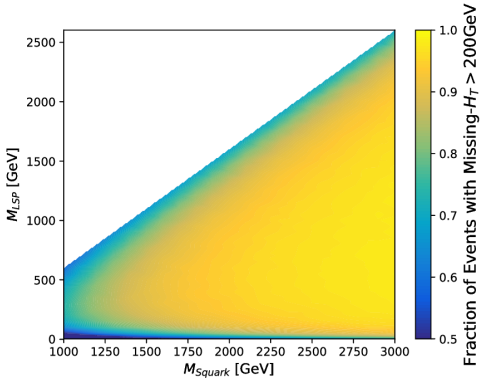
(a) BM1-type Mass Scan.



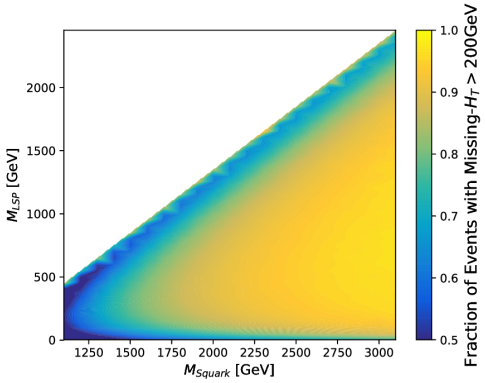
(b) BM2-type Mass Scan.



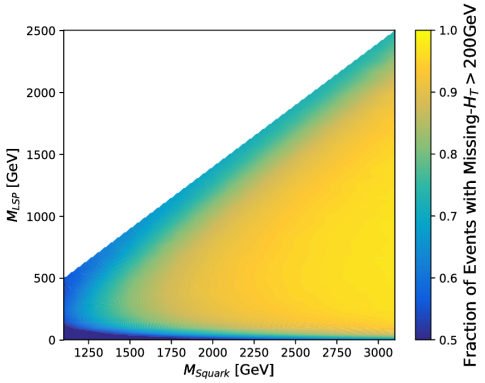
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.



(e) BM5-type Mass Scan.



(f) BM6-type Mass Scan.

Figure 6.4: Fraction of events with $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV for the BM1-BM6-type mass scans.

that we are left with heavy LSPs with very little kinetic energy, and thus low $H_{\mathrm{T}}^{\mathrm{miss}}$.

6.4.3 Number of hadronic jets

Typical SUSY decay cascades often involve a large number of hadronic jets being produced. Similarly, the NMSSM scenarios considered in this thesis feature two such cascades, each ending in a Higgs boson decay, with some additionally including stop or sbottom-type squarks whose decays result in even more jets.

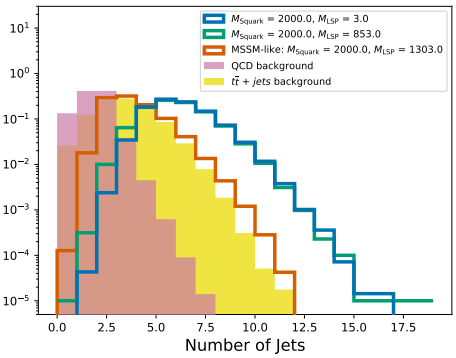
Figure 6.5 shows normalised distributions of the number of hadronic jets with $p_{\mathrm{T}}>$ 40GeV and $|\eta| < 2.4$ for example signal mass points in comparison with SM background processes.

Considering the fraction of events containing at least six hadronic jets, it can be seen in Figure 6.6 that whilst typically around 50% or more of the signal events pass this selection, the three scans where the gluino is lighter than the squark have a much higher efficiency with respect to this cut.

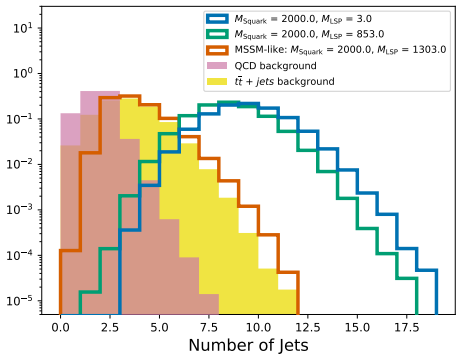
The primary reason for this behaviour lies in the decay products of the sparticles in the various mass hierarchies. The decay cascade for the BM2, BM4 and BM6-type scans involve squarks decaying to a gluino and a quark, where the gluino decays into two quarks and a neutralino. Conversely, scenarios where the squark is lighter than the gluino involve the gluino decaying into a squark and a quark, with the squark decaying into a neutralino and only one quark.

This three-body gluino decay means that each cascade, of which these scenarios include two, produces an extra quark, thus increasing the expected number of hadronic jets per event in the detector.

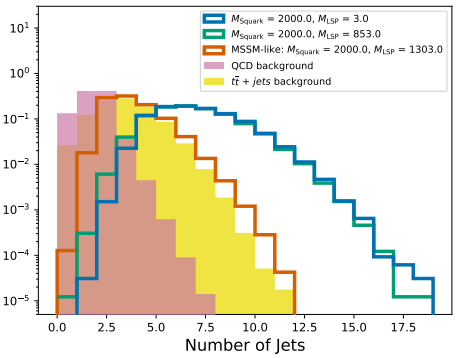
Additionally in each mass scan the fraction of events passing this selection is generally highest for a lighter LSP, dropping considerably as the LSP mass approaches the squark/gluino mass. This drop in efficiency for heavy LSP is due to the small mass gaps in the decay cascades meaning softer hadronic jets, such that the p_{T} of



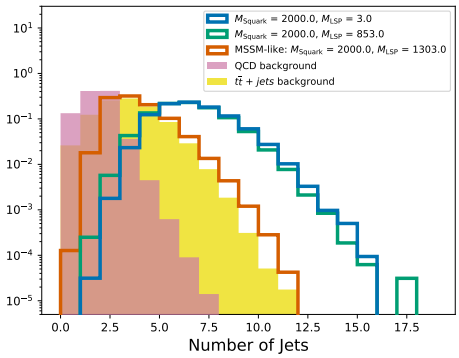
(a) BM1-type Mass Scan.



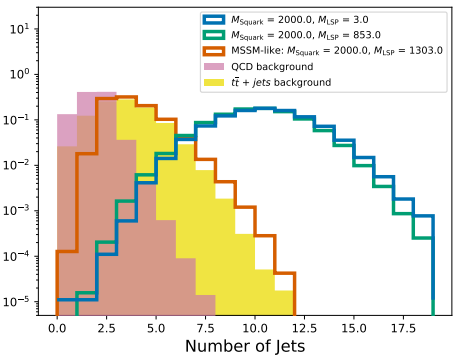
(b) BM2-type Mass Scan.



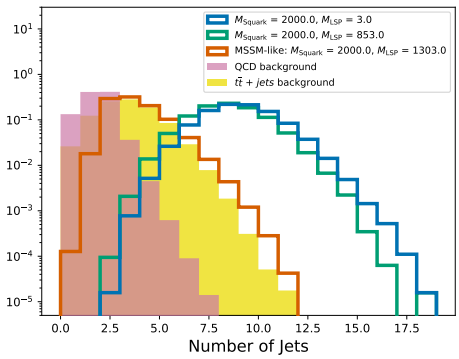
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.

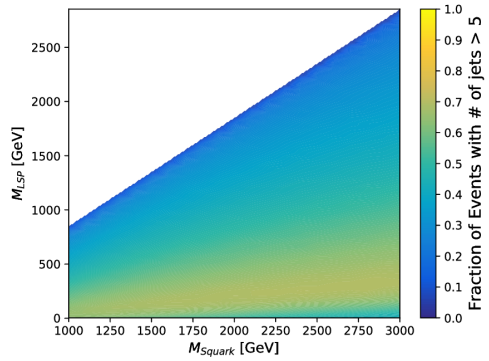


(e) BM5-type Mass Scan.

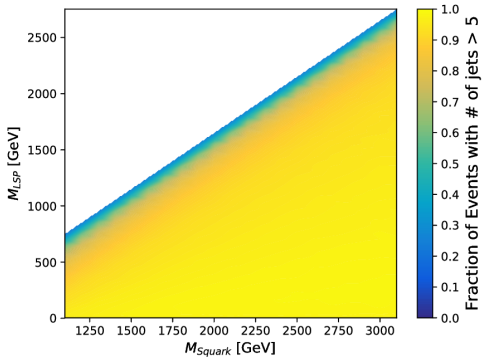


(f) BM6-type Mass Scan.

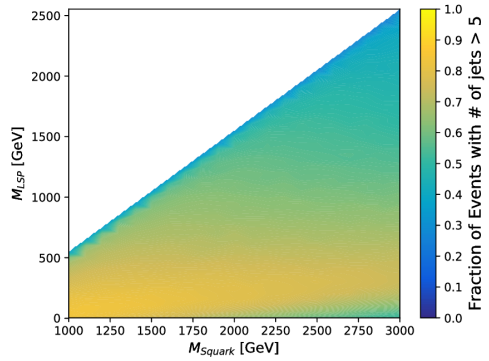
Figure 6.5: Number of hadronic jets for low and mid-range M_{LSP} near the observed limit in the BM1-type scan, compared with QCD and $t\bar{t}$ background processes and an MSSM-like scenario with a light LSP.



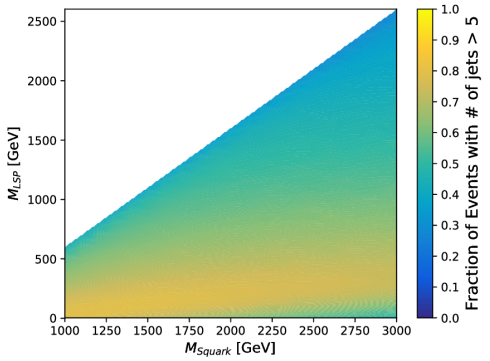
(a) BM1-type Mass Scan.



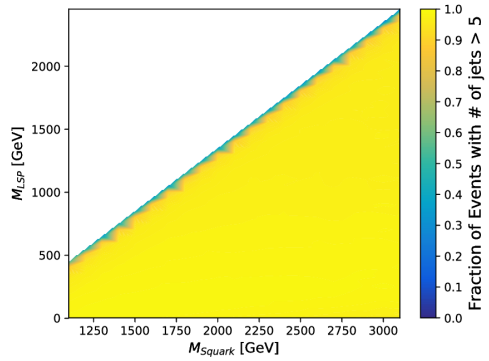
(b) BM2-type Mass Scan.



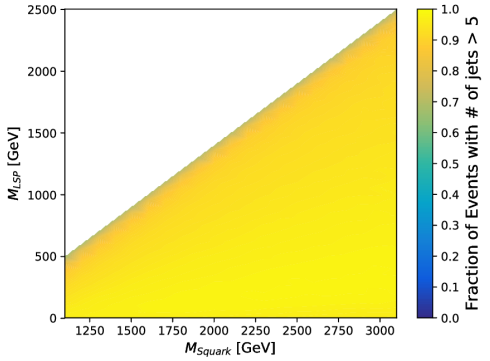
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.



(e) BM5-type Mass Scan.



(f) BM6-type Mass Scan.

Figure 6.6: Fraction of events with total number of hadronic jets $>$ 5 for the BM1-BM6-type mass scans.

some of these jets will fall below the 40 GeV minimum threshold and so will not be considered.

6.4.4 Number of *b*-tagged hadronic jets

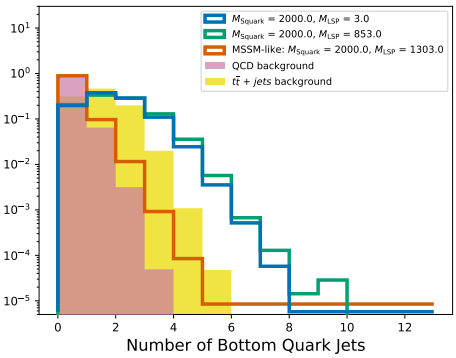
In this NMSSM scenario where the LSP is a singlino, two SM-like Higgs bosons will be produced. The BRs for SM-like Higgs boson decay dictate that the most likely decay is that to a bottom quark-antiquark pair, therefore it is expected that a large number of the hadronic jets in each event will be *tagged* as being a bottom quark, or *b-tagged*.

Not all *b*-tagged jets are necessarily hadronic jets containing bottom quarks, nor will all bottom quarks form *b*-tagged jets. However, the efficiency of correctly *b*-tagging a bottom quark is around 70%, whereas the likelihood of accidentally *b*-tagging a lighter flavour jet is only 1% or so. As such the average number of *b*-tagged jets for each of the example mass points in each mass scan is quite large, as shown in Figure 6.7.

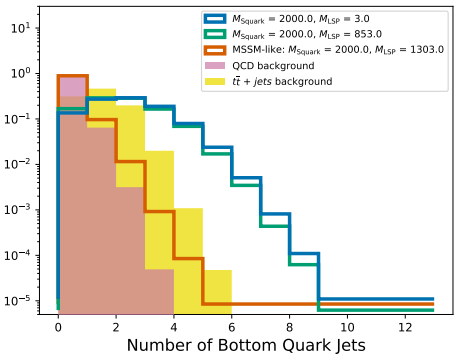
Considering the efficiency of the binning imposed, taking the fraction of events with at least two *b*-tagged hadronic jets, it may be noted that in general at least around half of the events in these signal points contain at least two *b*-tagged hadronic jets, as shown in Figure 6.8. However it is clear that in the BM5 and BM6-type scans this efficiency increases to almost 100%, shown also by the large number of such jets in the example mass points shown for these scans in Figures 6.7e and 6.7f.

The high *b*-tag multiplicities in these scans stem from the decay cascades, shown in Figure 2.6. In the case of the BM5 scan, up to four top quarks are produced, whose decays may lead to the production of bottom quarks. More simply, in the case of the BM6-type scan, it is possible to obtain up to four bottom quarks without even considering the decay of each Higgs boson.

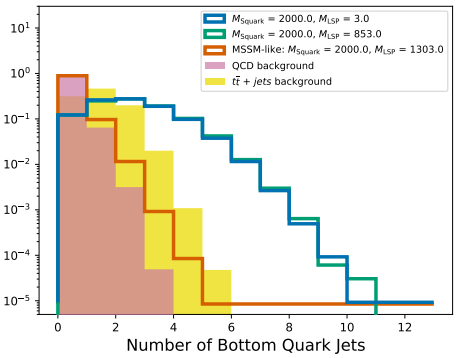
Additionally, it may be noted that whilst the correlations between sparticle masses and the fraction of events containing at least two *b*-tagged hadronic jets is not as clear as was seen in Figure 6.4 for $H_{\mathrm{T}}^{\mathrm{miss}}$, say, there is generally a decrease



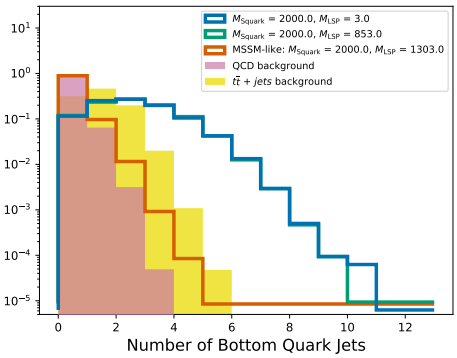
(a) BM1-type Mass Scan.



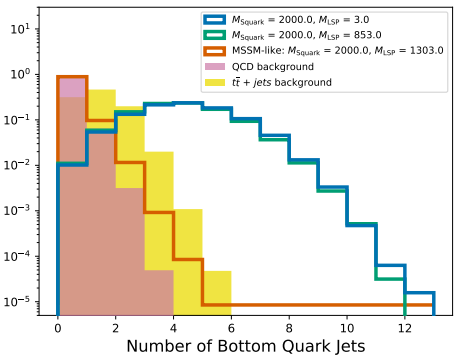
(b) BM2-type Mass Scan.



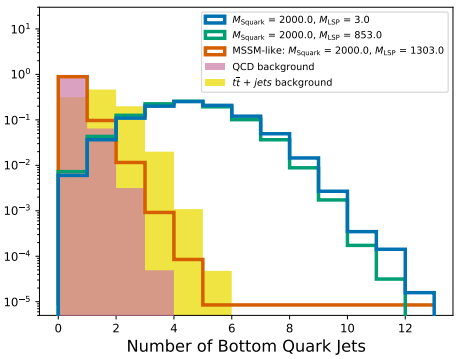
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.

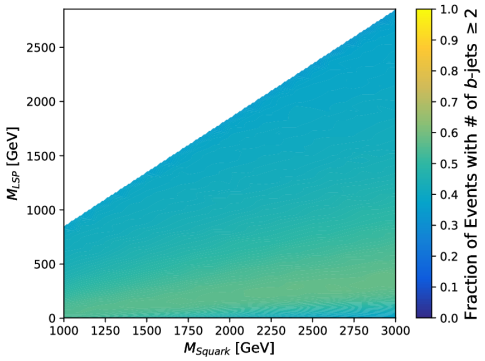


(e) BM5-type Mass Scan.

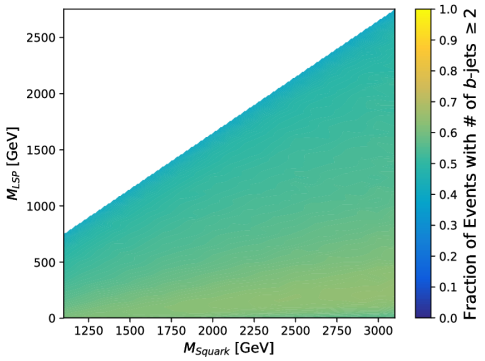


(f) BM6-type Mass Scan.

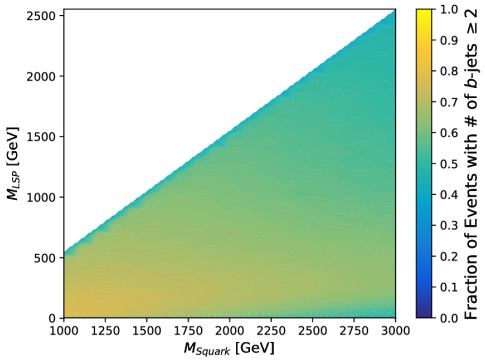
Figure 6.7: Number of b-tagged hadronic jets for low and mid-range M_{LSP} near the observed limit in the BM1-type scan, compared with QCD and $t\bar{t}$ background processes and an MSSM-like scenario with a light LSP.



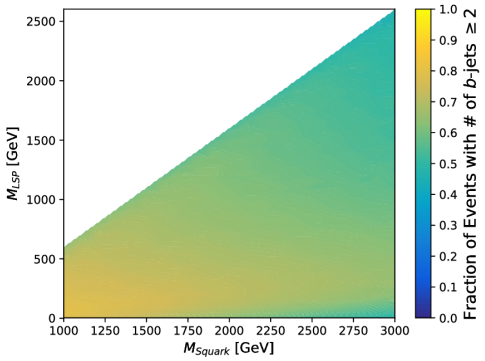
(a) BM1-type Mass Scan.



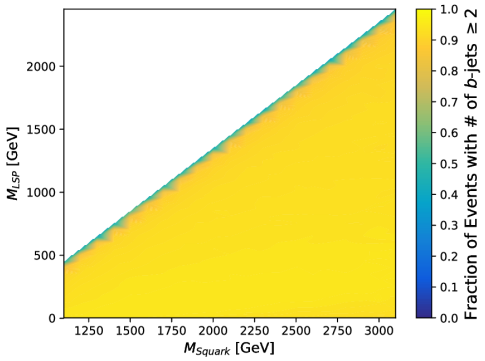
(b) BM2-type Mass Scan.



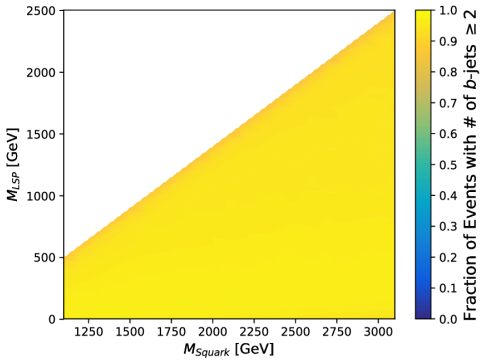
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.



(e) BM5-type Mass Scan.



(f) BM6-type Mass Scan.

Figure 6.8: Fraction of events with total number of b-tagged hadronic jets > 5 for the BM1-BM6 mass scans.

in efficiency as the LSP mass approaches the masses of the squark and gluino. Similarly to the case for the number of hadronic jets, this is likely since the jet p_{T} distribution is softer as the mass gaps in the SUSY cascade decrease, meaning some of the bottom quarks will not have enough transverse momentum required to pass the 40 GeV threshold necessary to be considered.

6.4.5 $\Delta\phi^*$

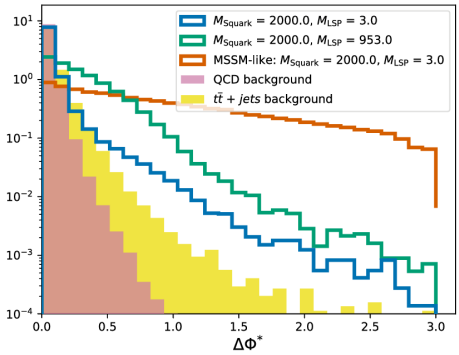
Defined in Section 5.4.3, $\Delta\phi^*$ is a variable used in order to reduce the background contribution from QCD multijet events. $\Delta\phi^*$ is designed in such a way that events with genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ or $H_{\mathrm{T}}^{\mathrm{miss}}$ would be expected to have large $\Delta\phi^*$ values whereas SM processes should generate small values, typically less than 0.5.

However, the NMSSM scenarios considered in this analysis do not generate many events containing large $\Delta\phi^*$, as shown in Figure 6.9. Considering the fraction of events with $\Delta\phi^* > 0.5$, shown in Figure 6.10, it becomes clear that a large fraction of events are rejected by the event selection, despite this variable being designed to reject QCD background events and to allow events with genuine $E_{\mathrm{T}}^{\mathrm{miss}}$.

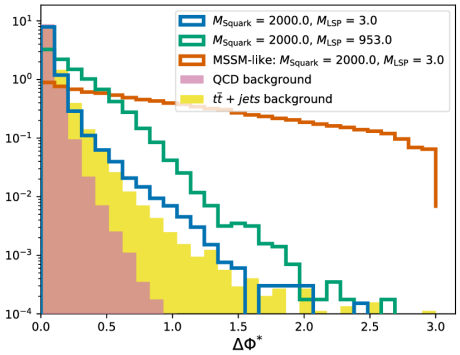
The relatively low $\Delta\phi^*$ values produced in these signal mass points indicates this particular background-reduction variable is not well-tuned to these NMSSM scenarios.

A likely reason for small $\Delta\phi^*$ lies in the large number of hadronic jets produced. Without of course assuming a uniform distribution in the direction of the hadronic jets, it would still be expected to become increasingly unlikely for the azimuthal angular separation between a hadronic jet and the $H_{\mathrm{T}}^{\mathrm{miss}}$ value computed without that jet to exceed the cut of 0.5 as the number of jets in each event increases. This is since for each sparticle decay the resulting quark will have momentum in a random direction in the sparticle rest frame. As such each jet will have an angular separation with respect to the other particles produced in the decay cascade. Since the angular separation in the transverse plane between any two objects cannot exceed π this means that, typically, as the number of jets in an event increases, the average angular distance between the jets decreases.

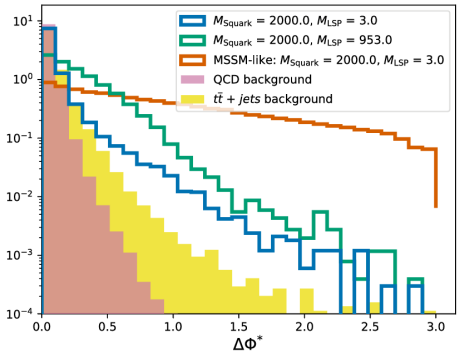
Accordingly, the regions of parameter space for which the LSP mass is close to that of the squarks and gluino display a larger fraction of events with $\Delta\phi^* > 0.5$, since the calculation of $\Delta\phi^*$ only involves hadronic jets with $p_{\mathrm{T}} > 40$ GeV, of which there are fewer in these regions.



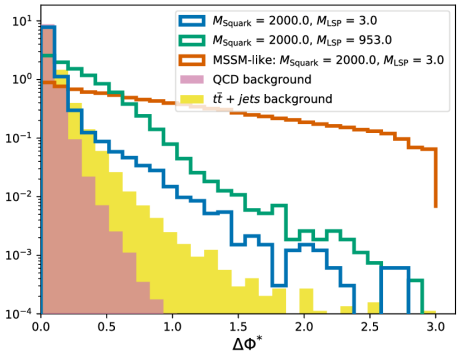
(a) BM1-type Mass Scan.



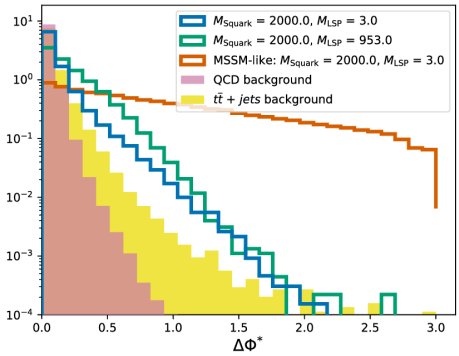
(b) BM2-type Mass Scan.



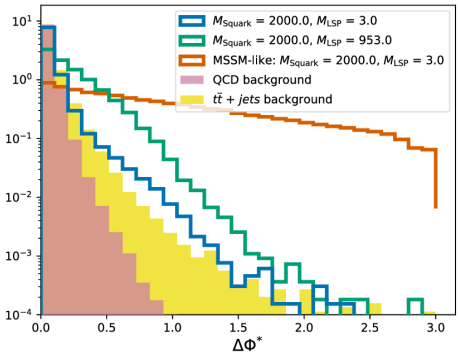
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.

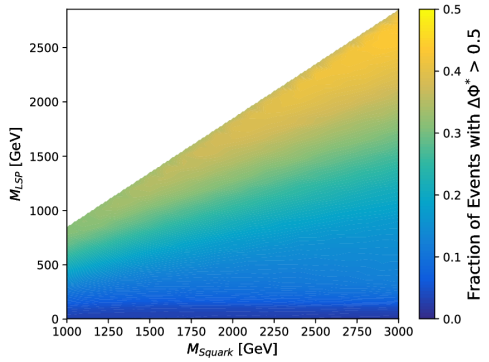


(e) BM5-type Mass Scan.

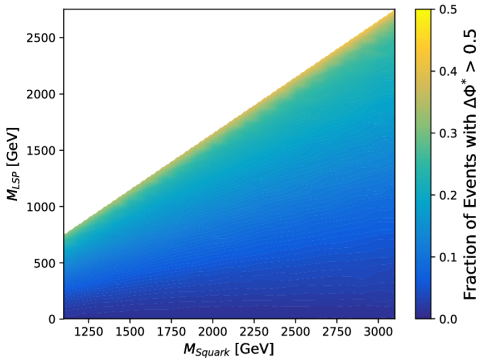


(f) BM6-type Mass Scan.

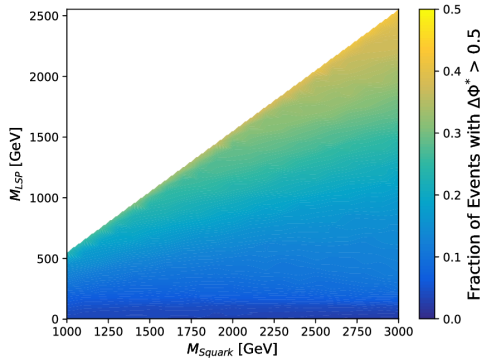
Figure 6.9: $\Delta\phi^*$ distributions for low and mid-range M_{LSP} near the observed limit in the BM1-type scan, compared with QCD and $t\bar{t}$ background processes and an MSSM-like scenario with a light LSP, where the selection requirement in this analysis is $\Delta\phi^* > 0.5$.



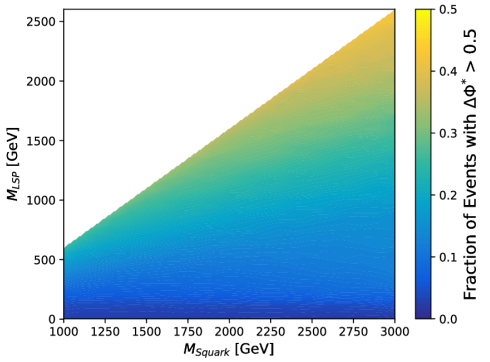
(a) BM1-type Mass Scan.



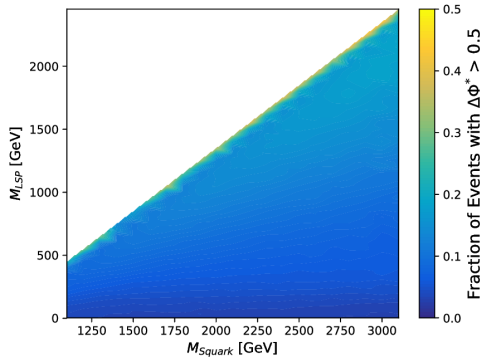
(b) BM2-type Mass Scan.



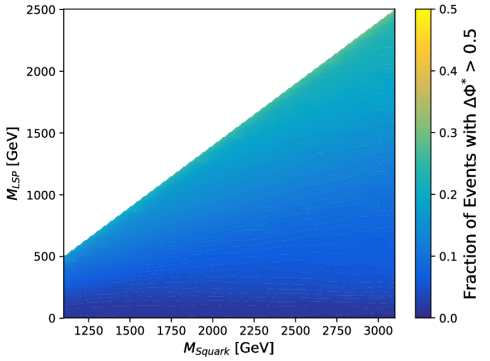
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.



(e) BM5-type Mass Scan.



(f) BM6-type Mass Scan.

Figure 6.10: Fraction of events with $\Delta\phi^* > 0.5$ for the BM1-BM6-type mass scans.

Given the very large average number of jets per event for the signal MC samples, especially those with a very light LSP, it would be expected that a $\Delta\phi^* > 0.5$ cut would indeed reject as much as 90% of the signal events in these regions of parameter space.

6.4.6 Angular separation between bottom quark jets from Higgs boson decays

One such quantity which characterises the BPs in this signal model is the angular separation ΔR between the bottom quarks from each Higgs boson decay, at MC generation ("truth") level. The bottom quark jets considered in [15] are standard AK4 jets, formed using the anti-k_{T} algorithm with a jet cone radius $R = 0.4$. If two such jets from a boosted Higgs boson are close enough that their angular separation ΔR is less than the jet cone radius then it might only be possible to resolve one *fat* bottom quark jet.

In order to examine the angular separation of the bottom quark jets stemming from the Higgs bosons, the MC truth information is examined. This contains the four-momenta of all individual particles in the event, rather than the emulated detector measurement of the hadronised jets, allowing for measurements of angular separation smaller than the jet cone radius of 0.4.

Considering the BM1-BM6-type mass scans it may be seen in Figure 6.11 how for example a heavy squark and a light LSP will correspond to a more boosted Higgs boson, thus decreasing the mean ΔR value. As the LSP mass increases, however, the bottom quark jets become more separated, since the LSP momentum increases relative to that of the associated Higgs boson, resulting in a more boosted LSP and a less boosted Higgs boson.

We note that the ΔR distributions in Figure 6.11 feature a kink around $\Delta R \approx \pi$. This is since $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ where $\Delta\eta$ is in the range $[0, 5]$ whereas $\Delta\phi$ takes values in the smaller range $[0, \pi]$.

Thus, in the extreme light LSP limit, it is expected that as much as 90% of the bottom quarks stemming from the decay of each Higgs boson will overlap to the extent that resolving and b-tagging both hadronic jets from each Higgs boson will become very difficult, as $\Delta R < 0.4$.

Here, it may be noted that the behaviour is essentially the same for each of these six mass scans, since in all cases heavy quarks combined with a light LSP leads to a boosted Higgs boson and, as such, small angular separation between the bottom quark-antiquark pair.

6.4.7 Validation of cut and count analysis tools

In order to check that one may rely on the signal event yields calculated by the software used to implement the event selection and indeed that the estimation of the systematic uncertainty is appropriate, it is important to compare these yields with those in [15]. However of course the experimental analysis in question does not feature an NMSSM low-$E_{\mathrm{T}}^{\mathrm{miss}}$ scenario such as those under consideration in this work, so a reference benchmark model is chosen, $T1bbbb$ [100].

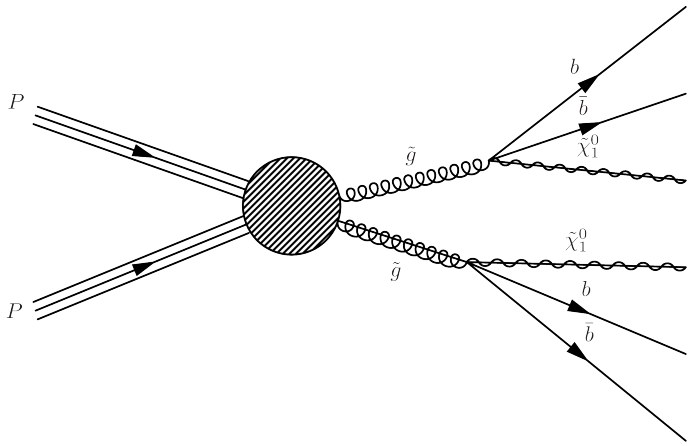
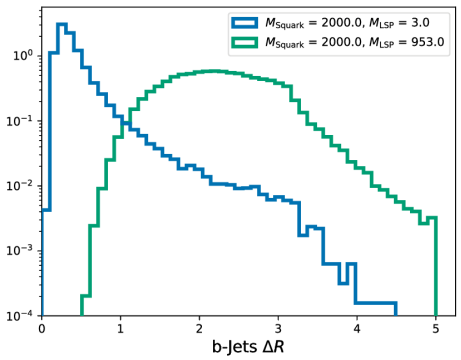
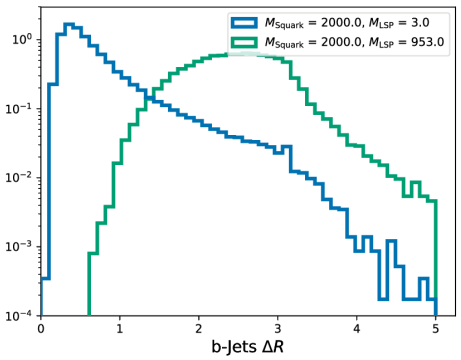


Figure 6.12: Feynman diagram showing gluino pair production and decay in the $T1bbbb$ benchmark model.

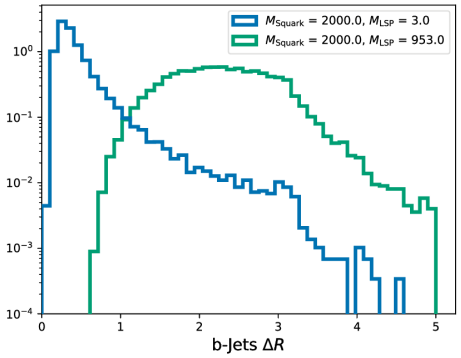
$T1bbbb$ is a simplified supersymmetric model whereby pair-produced gluinos each undergo a three-body decay into a bottom quark-antiquark pair and an LSP neutralino, as shown in Figure 6.12. In this example the gluino has a mass of 1900 GeV



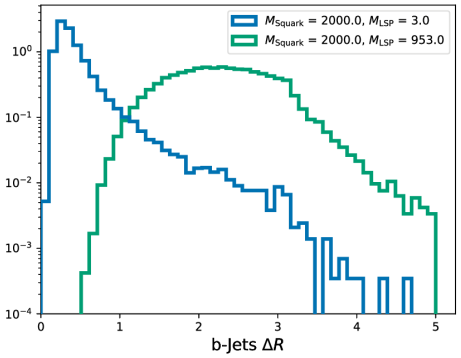
(a) BM1-type Mass Scan.



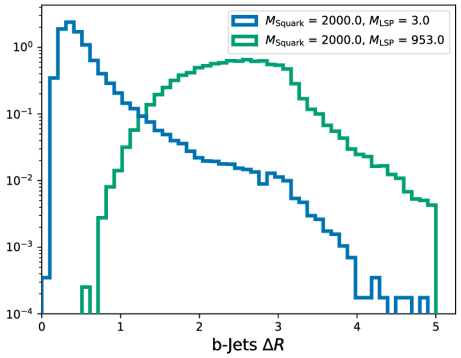
(b) BM2-type Mass Scan.



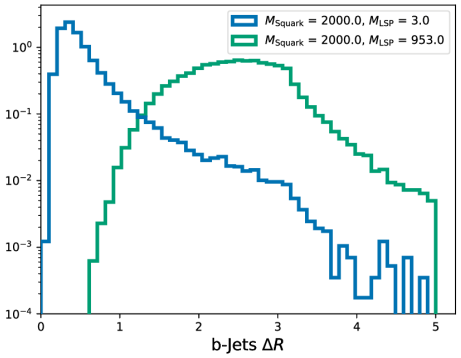
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.



(e) BM5-type Mass Scan.



(f) BM6-type Mass Scan.

Figure 6.11: ΔR distributions for low and mid-range M_{LSP} near the observed limit in the BM1-BM6-type mass scans.

Event Selection	Benchmark Model T1*bbbb*: $M_{\tilde{q}} = 1900\,\text{GeV}$, $M_{\chi_1^0} = 100\,\text{GeV}$	
	Efficiencies from Ref [15]	Delphes Implementation
Before selection	100.0	100.0
Muon, electron, photon vetoes	99.4	98.2
$p_T^{j_1} > 100$ GeV	98.7	98.1
$0.1 < f_{h^\pm}^{j_1} < 0.95$	93.9	98.1
$H_\text{T} > 200$ GeV	93.9	98.1
$H_\text{T}^\text{miss} > 200$ GeV	88.5	92.2
Event veto for forward jets ($\lvert\eta\rvert > 2.4$)	69.9	74.4
$H_\text{T}^\text{miss}/E_\text{T}^\text{miss} < 1.25$	69.3	73.7
n_jet- and H_T-dependent α_T thresholds	69.2	73.7
$\Delta\phi^*_\text{min} > 0.5$	25.1	23.7

Table 6.5: Cumulative percentages of events passing the event selections compared with those from Ref [15] for a standard reference benchmark model, *T*1*bbbb* [100].

and the LSP 100 GeV. In addition to the pair produced gluinos, up to two hard jets are considered at parton level in the event generation, as was done with the NMSSM signal mass points.

As shown in Table 6.5 it is clear that the respective efficiencies of each of the event selections are all within a few percent of those taken directly from [15] for the same example benchmark model. Therefore, in order to take into account other sources of uncertainty and the results from the validation exercise for the Prospino signal production cross-section values, shown in Section 6.2.1, we assign an overall systematic uncertainty value of 25%.

6.5 Results

6.5.1 CL_S Method

In order to place limits on new physics processes, an assessment is performed comparing this signal-plus-background hypothesis with the null hypothesis; that no new physics contribution exists in the data. Say we have some observable x

which allows for the construction of a histogram $\vec{n} = (n_1, ..., n_N)$. We may then write the expectation value for the number of events in bin i of $\vec{n}$ as:

$$E(n_i) = \mu s_i + b_i, \tag{6.1}$$

where s_i and b_i are the predicted signal and background yields in bin i respectively [101]. Here μ represents the signal strength, whereby $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ the nominal signal hypothesis.

We then have a likelihood function $\mathcal{L}$ across the N bins, defined as the product of Poisson distributions as follows:

$$\mathcal{L}(\mu, \vec{\theta}) = \Pi_{i=1}^{N} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} \cdot e^{-(\mu s_i + b_i)}, \tag{6.2}$$

where $\vec{\theta} = (\vec{\theta}_s, \vec{\theta}_b)$ represents nuisance parameters which characterise the probability density functions describing the distributions of s_i and b_i respectively [101].

The profile likelihood ratio is then defined as

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})}, \tag{6.3}$$

where $\hat{\hat{\vec{\theta}}}$ maximises $\mathcal{L}$ for a given value of μ, whereas $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the μ and $\vec{\theta}$ maximise $\mathcal{L}$ unconditionally.

For the purpose of determining an upper limit on the strength parameter μ, we define a test statistic q_μ as follows [101]:

$$q_\mu = \begin{cases} -2\ln\lambda(\mu), & \hat{\mu} \leq \mu, \\ 0, & \hat{\mu} > \mu. \end{cases} \tag{6.4}$$

Here the reason for setting q_μ to zero in the case where $\hat{\mu} > \mu$ is that in the case where an excess in data over the background estimation allowed for a signal

strength greater than one, one would not regard this as being less compatible with the signal hypothesis compared with the background-only hypothesis. Therefore in these cases we would not consider $\hat{\mu} > \mu$ to be a rejection of the test. In the case of the NMSSM scenarios under investigation in this thesis we consider only the case where both Higgs bosons decay into a bottom quark-antiquark pair, with other decay channels not being expected to contribute to the final expected yields due to the H_{T}, N_{jet} and $N_{b\text{-jet}}$ requirements. As such, no interference effects are considered between signal and background processes.

Now we define the $\mathrm{CL_S}$ method [102]. Having obtained an observed value for the test statistic, denoted $q_{\mu,\mathrm{obs}}$, we consider the probability of finding some data set with which the nominal signal hypothesis is at least as compatible:

$$CL_{S+B} = p_{S+B} = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f_{S+B}(q_\mu|\mu)dq_\mu, \tag{6.5}$$

where f_{S+B} represents the probability density function describing the distribution of q_μ for the nominal signal hypothesis.

Similarly, the degree to which the background-only hypothesis (where $\mu = 0$) represents the data, denoted p-value, may be written as follows:

$$p_B = 1 - CL_B = \int_{-\infty}^{q_{\mu,\mathrm{obs}}} f_B(q_\mu|0)dq_\mu. \tag{6.6}$$

Finally, the $\mathrm{CL_S}$ parameter gives the compatibility between a given signal model and the data:

$$\frac{CL_{S+B}}{CL_B} = \mathrm{CL_S} = \frac{p_{S+B}}{1 - p_B}. \tag{6.7}$$

In the case where $\mathrm{CL_S} \geq 0.05$, a signal model is considered to be excluded at 95 % confidence level (CL) [102].

6.5.2 Signal Model Interpretations

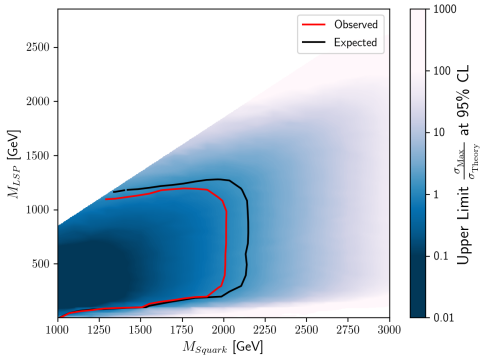
Using the asymptotic CLs method described in Section 6.5.1, a fit across the signal, background and data yields in the nine measurement bins described in Table 6.4 is performed in order to determine an upper limit on the strength parameter μ, defined as the upper limit placed on the fraction of the signal cross-section, at 95% CL. As such, if for a particular point in parameter space $\mu < 1$ this point may be excluded, corresponding to the case where $\mathrm{CL_S} \geq 0.05$, whereas if $\mu > 1$ the point may not be ruled out.

These limits are calculated given the data and background yields as well as background uncertainties in Table 6.4, with the systematic uncertainty on the signal yields assumed as 25%.

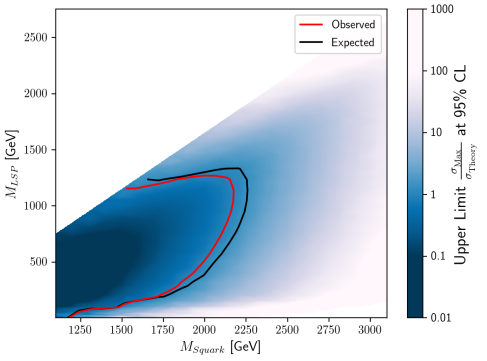
Here we present the contour limit plots for the six scan types performed within the NMSSM. The X- and Y-axes represent the masses of the squark and LSP, respectively, while the colour scale shows the upper limit on the strength parameter, μ. The black contour at $\mu = 1$ then identifies the areas in parameter space inside which all mass points are ruled out, assuming the data yields match the background expectation exactly. This is known as the expected limit, and allows us to quantify any excess or deficit in the observation compared to what we would expect given the background-only hypothesis.

Similarly the observed limit, shown as a red contour, is the upper limit given the actual observed data yield and the background and signal estimations.

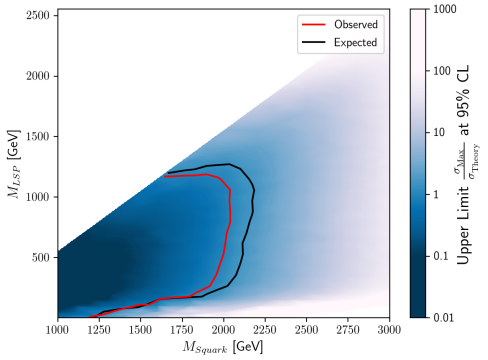
In Figure 6.13 it can be seen that both the observed and expected limits are far weaker for a very light LSP mass, with both contours bending to the left such that points with much lower squark masses are no longer excluded. As such it would appear the sensitivity of the analysis in [15] decreases dramatically in the limit of a light LSP in these NMSSM scenarios, with the lower bound on the squark mass decreasing from around 2 TeV to as little as 1 TeV in some cases, as shown in Table 6.6.



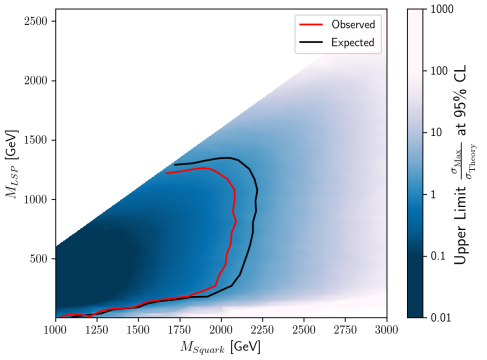
(a) BM1-type Mass Scan.



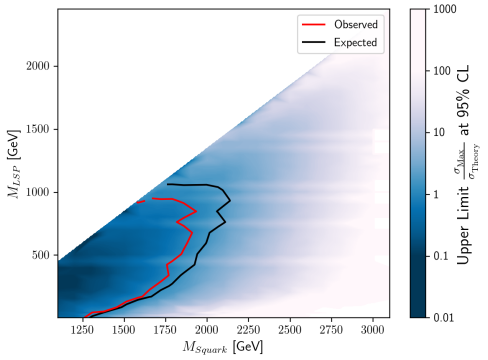
(b) BM2-type Mass Scan.



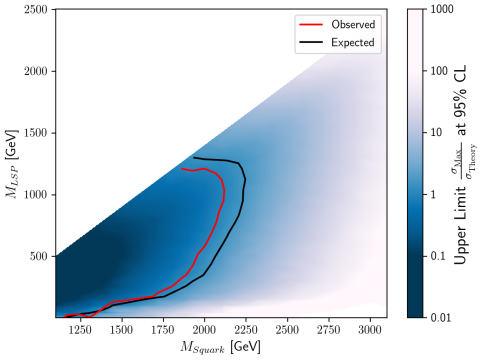
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.



(e) BM5-type Mass Scan.



(f) BM6-type Mass Scan.

Figure 6.13: Observed and expected limits for the BM1-BM6-type mass scans. The X- and Y-axes represent the squark and LSP masses, respectively, whilst the colour scale represents the upper limit on the strength parameter μ.

Scan	BM1	BM2	BM3	BM4	BM5	BM6
$M_{\tilde{q},\,\mathrm{min}}$ [GeV/c^2]	1000	1200	1250	1000	1250	1200
$M_{\tilde{g},\,\mathrm{min}}$ [GeV/c^2]	1010	1000	1260	1010	1050	1000

Table 6.6: Approximate lower bounds on the squark mass and corresponding gluino mass at 95% CL for a 3 GeV LSP.

Additionally whilst the red observed limit contour is generally further to the left than the black expected limit contour, indicating a slight excess in some of the data yields compared with the background estimation, the agreement between the two limits is reasonably strong.

These weaker lower bounds on the squark masses for the lightest LSP mass of 3 GeV are summarised in Table 6.6.

It may be noted that the lower bounds on the squark and gluino masses are considerably weaker for these light-LSP, low-$E_{\mathrm{T}}^{\mathrm{miss}}$ scenarios compared with typical MSSM-like simplified models [15], with a comparison shown in Section 6.5.3.

As the LSP mass is increased above around 100 GeV the converse becomes true, with the limits being more harsh for these NMSSM scenarios than for typical MSSM-like simplified models [15], with an example MSSM-like scenario shown for comparison in Section 6.5.3. This is expected due to the larger $H_{\mathrm{T}}^{\mathrm{miss}}$ in these heavier LSP regions.

However, as the LSP mass is increased closer towards the masses of the squarks and gluinos, the sensitivity appears once again to decrease for heavier neutralinos. Unlike the light LSP region, however, this lack of sensitivity for heavy LSP likely arises from the high H_{T} cut, since few events in this region pass this cut as shown in figure 6.2a. Thus, in order to explore this area of mass space a wider H_{T} range would be required than is considered in this particular analysis.

Similar experimental limits can be placed on the other types of scan. We recall here that the BM2-type scan has the gluino mass 200 GeV lower than the squark mass, rather than 10 GeV higher, and the BM3/BM4- and BM5/BM6-type scans are the same as the BM1- and BM2-types, respectively, but with the appropriate

stop or sbottom squark masses 250 GeV lighter than the squark/gluino, rather than being decoupled.

The observed and expected limits for these remaining mass scans exhibit a similar behaviour, that is, the cross-section appears to dominate the sensitivity for points with mid-range LSP mass, where the contours are closer to vertical. However, in all cases the sensitivity for regions with lower LSP masses and featuring high H_{T} and low $H_{\mathrm{T}}^{\mathrm{miss}}$, to the latter of which the analysis in [15] is not optimised, decreases dramatically.

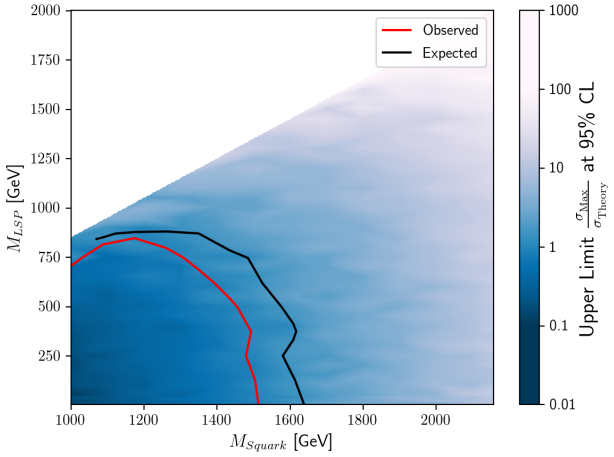
6.5.3 MSSM-like scenarios with light LSP

The main feature of the light LSP and low $E_{\mathrm{T}}^{\mathrm{miss}}$ scenarios under consideration relies on the LSP being singlino. In this case it is possible for the decay cascades to end exclusively in an NLSP decaying to an LSP and a Higgs boson, which is not the case in the MSSM whereby the squarks and gluinos may decay directly into hadronic jets and an LSP, giving rise to large $E_{\mathrm{T}}^{\mathrm{miss}}$ even in the case where $M_{\tilde{\chi}_1^0}$ is only a few GeV.

For comparison, Figure 6.14 shows the sensitivity to the simplified MSSM-like scenario seen previously. Applying the same event selections as in [15], the observed and expected limits are calculated at 95% CL, akin to the limits in Figure 6.13.

In the limit plots in Figures 6.13 and 6.14 it may be observed that the experimental sensitivity to this simplified MSSM-like model is in fact strongest for the lightest LSP mass of 3 GeV, contrary to the NMSSM-specific low-$E_{\mathrm{T}}^{\mathrm{miss}}$ mass scenarios.

Conversely, it is also clear that the overall sensitivity for the regions of parameter space in which $M_{\tilde{\chi}_1^0} > 200$ GeV is weaker for this model compared with the NMSSM scenarios considered. Since the decay cascade is truncated and there are no Higgs boson decays in this model, the expected numbers of hadronic jets and b-tagged jets per event are lower. Therefore, it is unlikely for as many events to contain more than five hadronic jets or as many as two b-tagged jets, and thus they will not



(a) BM1-type Mass Scan.

Figure 6.14: Observed and expected limits for an MSSM-like scenario demonstrating the higher sensitivity to regions with low LSP mass.

pass the full kinematic event selections. However, exploiting the full 254 bins in Ref [15] would be expected to increase the sensitivity to this MSSM-like scenario.

6.6 Conclusions of Reinterpretation of Jets+$\boldsymbol{E}_{\mathbf{T}}^{\mathbf{miss}}$ Analysis

Fairly strong limits of around 2 TeV have been placed on the squark/gluino masses for a singlino-like LSP in the NMSSM for a LSP mass above $\sim$ 100 GeV. However, below this mass the limits weaken by a considerable amount in all cases, as shown in table 6.6, with limits for some scenarios decreasing by as much as 1 TeV.

The eight original BPs in Ref [37], all featuring a light 3 GeV LSP, despite having a large direct production cross-section, are still on or around the limit of sensitivity of the analysis in Ref [15] with 35.9 fb^{-1} data from the CMS detector at the LHC, thus they cannot be completely excluded at this stage.

It is clear that these light LSP and low $E_{\mathrm{T}}^{\mathrm{miss}}$ scenarios present further challenges for jets+$E_{\mathrm{T}}^{\mathrm{miss}}$ based searches akin to Ref [15] at the LHC. In order to develop a search for these stealthy scenarios one might wish to access regions of low $H_{\mathrm{T}}^{\mathrm{miss}}$, however, this would require careful techniques so as to not allow yields from background processes to dominate.

As we will see in the following chapters in this thesis, the use of a double-b-tagger for the purpose of identifying events containing boosted Higgs bosons decaying to bottom quark-antiquark pairs allows for improved sensitivity to these NMSSM scenarios at the LHC.

6.7 Summary

In this chapter we have explored the ability of a general purpose jets+$E_{\mathrm{T}}^{\mathrm{miss}}$ SUSY search [15] to probe the parameter space of the NMSSM scenarios considered in this thesis. It was found that the observed and expected limits placed on the squark and gluino masses in the case of a heavier LSP is around 2 TeV. However, for a light LSP with mass of just a few GeV, this lower bound on the squark and gluino masses drops considerably, by as much as 1 TeV, for all six of the discrete mass scans considered.

Therefore in the case of a very light LSP we see that these low-$E_{\mathrm{T}}^{\mathrm{miss}}$ NMSSM scenarios would have escaped detection by an analysis such as that considered in this chapter, and so a novel approach is required. In the following chapter we detail and expand upon an analysis which focuses on SUSY cascades featuring boosted Higgs bosons and small $E_{\mathrm{T}}^{\mathrm{miss}}$, targeting boosted $H \to b\bar{b}$ decays with the use of a double-b-tag discriminator.

7 | CMS search for NMSSM Signatures with Low Missing Transverse Energy

7.1 Analysis Overview

As has been shown in Chapter 6, the otherwise strong sensitivity towards these NMSSM scenarios demonstrated by jets+$E_{\mathrm{T}}^{\mathrm{miss}}$ searches decreases considerably for regions where $M_{\tilde{\chi}_1^0} \lesssim 300$ GeV, highlighted yellow in Figure 7.1. As $M_{\tilde{\chi}_1^0} \to 0$ it is expected that the Higgs bosons will inherit an increasing fraction of the NLSP momenta, since the NLSP decays with 100% branching ratio into an LSP and a Higgs Boson, resulting in less $E_{\mathrm{T}}^{\mathrm{miss}}$ and highly boosted bottom quark-antiquark pairs stemming from the Higgs boson decays. In the case of an LSP with mass greater than that of the SM-like Higgs boson, 125 GeV, the $H_{\mathrm{T}}^{\mathrm{miss}}$ is expected to be higher. However, in the case where $M_{\tilde{\chi}_1^0} < 200$ GeV, the Higgs boson will still inherit a non-negligible fraction of the boosted NLSP momentum, and so the $H_{\mathrm{T}}^{\mathrm{miss}}$ will not be as high as is more typical for a simplified MSSM-like model.

In order to better target this region of parameter space we interpret the NMSSM scenarios under investigation in this thesis in the context of a novel CMS analysis [2], the event selection of which is detailed in Section 7.2. This analysis searches for high-H_{T}, low-$E_{\mathrm{T}}^{\mathrm{miss}}$ NMSSM scenarios very similar to those under investigation in this thesis, but with the decay cascades featuring light scalar Higgs bosons decaying to bottom quark-antiquark pairs, rather than SM-like Higgs bosons.

The key, novel feature of this analysis is the use of a double-b-tag discriminator to identify boosted objects, such as Higgs bosons, decaying to a bottom quark-antiquark pair, contained within a single AK8 jet. Since in the case of a highly boosted NLSP we expect the bottom quark-antiquark pair produced by each Higgs

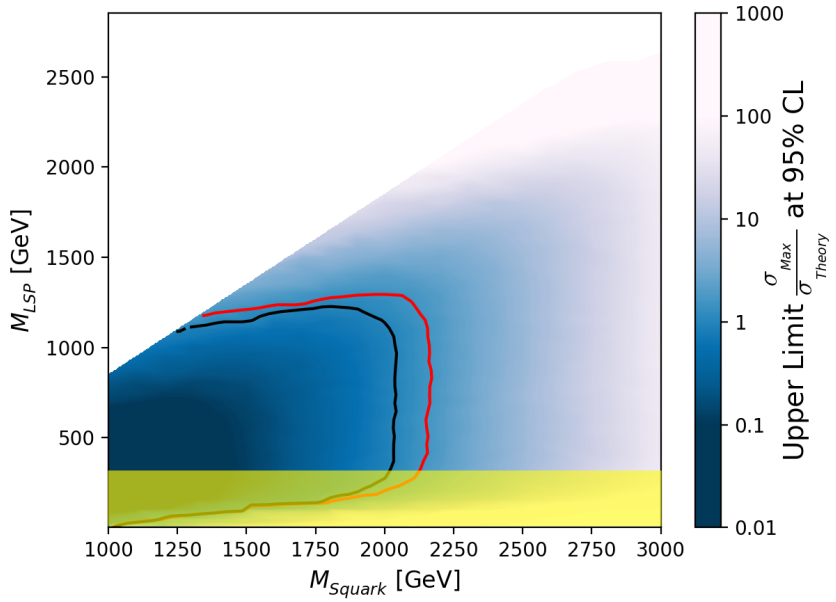


Figure 7.1: Limit plot for the BM1 mass scan from Chapter 6 with the region of interest, for which $M_{\tilde{\chi}_1^0} \lesssim 300$ GeV, highlighted yellow. Here the red and black contour lines represent the observed and expected limits at 95% CL respectively obtained by re-interpretation of the analysis in Ref [15].

boson decay to have very small angular separation, it is expected that forming a larger radius AK8 jet and examining the substructure, in order to assign such a double-b-tag, is well suited to the NMSSM scenarios under investigation in this thesis. As is detailed in Section 7.2 we then consider the reconstructed masses of the two jets with highest double-b-tag score in order to distinguish NMSSM events containing small $E_{\mathrm{T}}^{\mathrm{miss}}$ and boosted $H \to \mathrm{b\bar{b}}$ decays from SM background contributions.

Since the existing analysis only considers very light LSP masses of $\sim$ 1 GeV, producing small $E_{\mathrm{T}}^{\mathrm{miss}}$, we then modify this analysis by adding measurement bins for an additional variable, $H_{\mathrm{T}}^{\mathrm{miss}}$, so as to be better adapted to the particular NMSSM scenarios under investigation in this thesis as the LSP mass increases towards 300 GeV. This is done in order to gain sensitivity to the regions of parameter space to which the analysis in Chapter 6 was not optimised, highlighted in Figure 7.1.

This chapter thus consists of an overview and interpretation of Ref [2], to which the author contributed, in the context of the NMSSM scenarios under investigation in this thesis, as well as an $H_{\mathrm{T}}^{\mathrm{miss}}$-based extension to the event selection.

The dominant background contribution in this analysis is from QCD multijet events, which is estimated from data using control regions. Sub-dominant background processes such as $t\bar{t}$ production are predicted using MC simulated events, with checks carried out to ensure good agreement between observed and simulated events.

Finally the expected signal and background yields across the measurement bins are compared with the observed yields from data taken at the CMS detector in run years 2016 and 2017. Data collected in 2018 is not used for this analysis, owing to time constraints regarding the present lack of collaboration-approved double-b-tag scale factors for this most recent run year. The final results are marked "CMS Work In Progress", since the analysis work in this chapter has not yet undergone full collaboration approval, and is not currently in the public domain.

7.2 Event Selection

In this section we examine the baseline event selection in the analysis in Ref [2] and assess its suitability for the purposes of searching for the NMSSM scenarios under investigation in this thesis. Firstly we explore event selection requirements chosen in order to maximise the number of signal process events remaining, whilst reducing contributions from background processes as much as possible.

For some variables such as jet p_{T}, simple cuts are applied. However as is detailed later in this selection, measurement bins are constructed within certain variables, allowing for finer measurement of the shapes of the respective distributions, in order to better characterise and discriminate between the various signal and background processes. In the case of the distribution of a variable being divided into more than one measurement bin, the expected and observed event yields in neighbouring bins will be correlated, since a small difference in the measured quantity

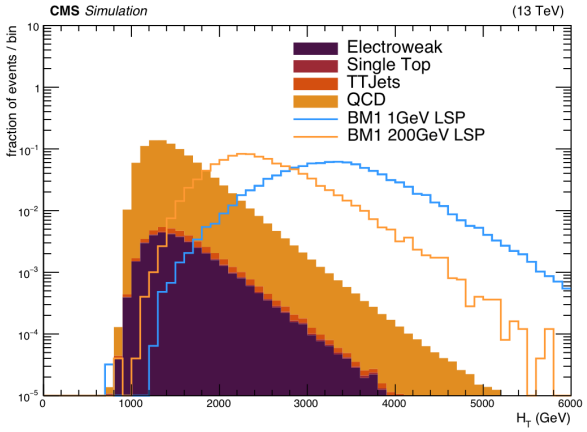
could mean that an event belonging to a different bin. For the sake of brevity, figures shown in this chapter represent a sub-set of the NMSSM mass hierarchies under investigation, with a full set of figures found in Appendix A.

7.2.1 Baseline Cuts

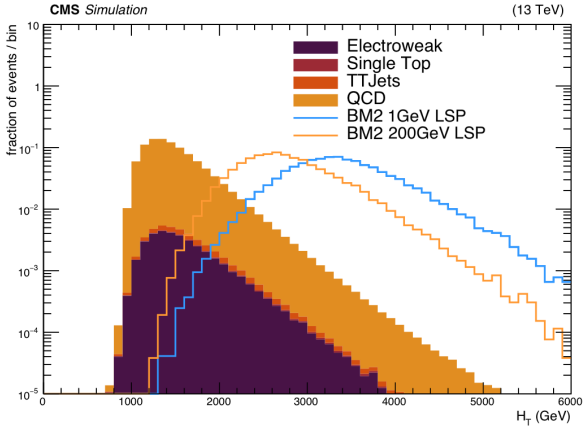
As detailed in Chapter 6, the NMSSM scenarios under consideration feature a large number of high-p_{T} hadronic jets, stemming from the gluino, squark and Higgs boson decays. Therefore, in order to reject a greater number of background events whilst having little to no impact on the overall signal efficiency we define the baseline requirements, i.e. those which are imposed across all measurement bins, based upon the AK4 and AK8 jets in each event as follows:

- At least two AK8 jets, allowing for the application of AK8 jet substructure techniques, with $|\eta| < 2.4$ (central jets).
- The two central AK8 jets with highest double-b-tag score must both have $p_{\mathrm{T}} > 300$ GeV.
- Double-b-tag score selection criteria as defined in Section 7.2.3.
- AK8 jet soft-drop mass selection criteria as defined in Section 7.2.4.
- At least one AK4 jet with $|\eta| < 3.0$, $p_{\mathrm{T}} > 300$ GeV and angular separation $\Delta R > 1.4$ between this jet and both of the selected AK8 jets defined above.
- $H_{\mathrm{T}} > 1500$ GeV.

The angular separation $\Delta R > 1.4$ between the selected AK8 jets and the leading AK4 jet is a conservative choice, chosen in order to ensure there is no overlap between particle flow candidates entering the reconstruction of any AK4 jet and those of either AK8 jet.



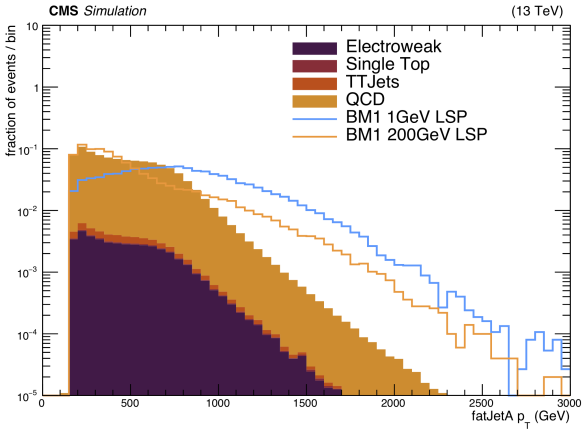
(a) BM1 Mass Scan.



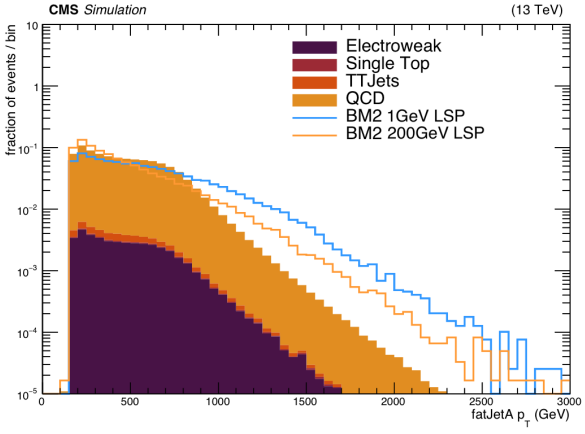
(b) BM2 Mass Scan.

Figure 7.2: Normalised scalar H_{T} distribution comparison between example signal mass points from the BM1 and BM2 mass scans and SM background processes.

The baseline H_{T} requirement of $H_{\mathrm{T}} > 1500$ GeV was chosen since it has very little effect on the signal spectra in these NMSSM scenarios for a wide range of squark and gluino masses, as well as LSP masses up to as high as 200 GeV, whilst reducing the contributions from background processes considerably, as shown in figure 7.2. Additionally, as is shown in figures 7.3 and 7.4 respectively, a similar effect is observed for the baseline cuts regarding the p_{T} of the AK8 double-b-tagged jets and the leading AK4 jet, with the mean of these distributions in excess of the respective threshold, since the light-LSP NMSSM scenarios considered in this thesis typically produce high-p_{T} hadronic jets, leading to high H_{T}.

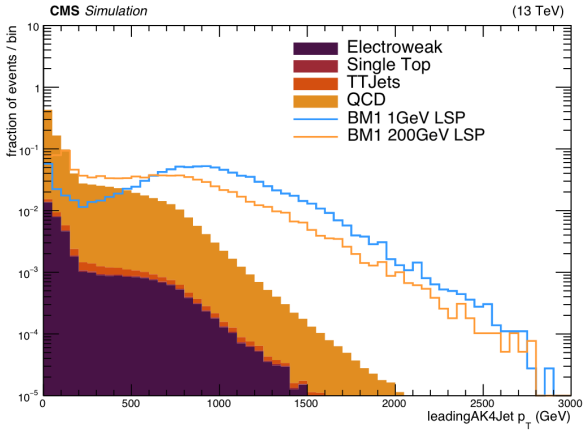


(a) BM1 Mass Scan.

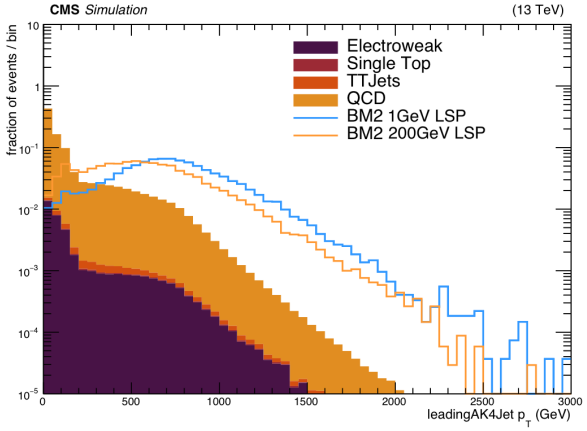


(b) BM2 Mass Scan.

Figure 7.3: Normalised AK8 jet p_{T} distribution comparison between example signal mass points from the BM1 and BM2 mass scans and SM background processes.



(a) BM1 Mass Scan.



(b) BM2 Mass Scan.

Figure 7.4: Normalised leading AK4 jet p_{T} distribution comparison between example signal mass points from the BM1 and BM2 mass scans and SM background processes.

7.2.2 H_{T} binning

As the squark, gluino and LSP masses vary, so does the mean H_{T} value. As is shown in figure 7.2, varying the NLSP and LSP masses has a notable effect on the H_{T} distribution, since a heavier LSP inherits a larger fraction of the NLSP momentum.

Therefore, in order to improve discrimination between signal+background and background-only expected yields, the H_{T} measurement is split into three bins:

$$H_{\mathrm{T}} \in [1500, 2500), [2500, 3500), [3500, \infty) \text{ GeV}$$

As is also shown in figure 7.2 the background process yields will be reduced considerably following this H_{T} requirement. However, owing to the large production cross-section values for processes such as QCD and $t\bar{t}$ further background reduction steps are required.

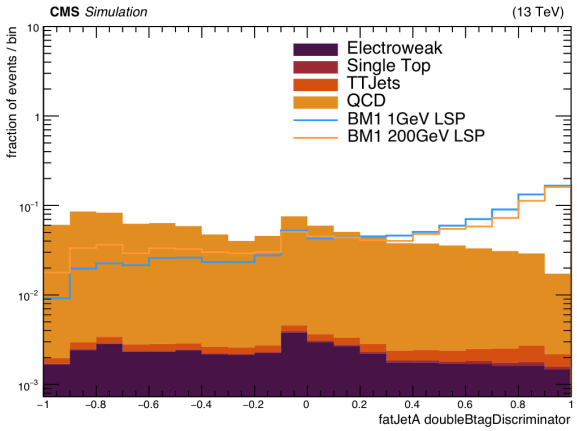
7.2.3 2D AK8 Double-*b*-tag Score Selection

Since the final state under consideration is that for which two Higgs bosons each decay into a boosted bottom quark-antiquark pair, too closely overlapping to allow for resolving the quarks as individual b-tagged jets, the focus instead is to construct a larger-radius jet cone around each $b\bar{b}$ pair. This jet, having twice the radius parameter of a regular AK4 jet, is then assigned a discriminator score indicating whether it has been identified as containing two bottom quarks stemming from the same object – in this case, a Higgs boson.

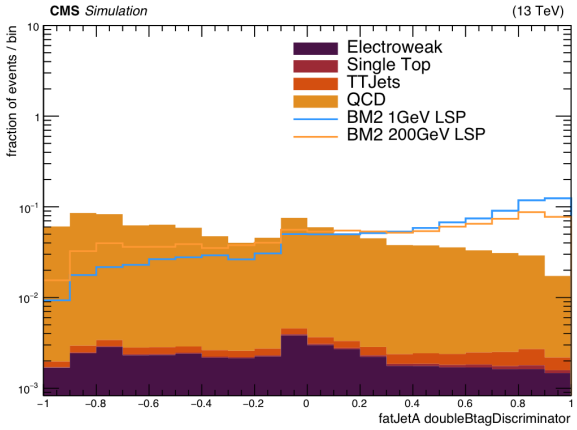
This discriminator score is driven by a Boosted Decision Tree (BDT) [90] which, as described in Section 5.1.3, uses 27 measurement variables to determine the output [103], between -1 and 1, representing the likelihood for a given jet to have originated from an object decaying into a bottom quark-antiquark pair.

Consider the two central AK8 jets, each having $p_T > 300$ GeV, with the highest double-b-tag discriminator score. We denote these jets A and B, with double-b-tag scores D_{bb}^{A} and D_{bb}^{B} respectively, in random order since the QCD background estimation method detailed in Section 7.5 requires the various 2D distributions to be unchanged when swapping jets A$\leftrightarrow$B. Figure 7.5 shows the distribution of the double-b-tag score of AK8 jet A for example signal mass points compared with SM background processes. It is clear here that these NMSSM scenarios generate

events containing AK8 jets with very high double-b-tag scores, owing to the highly boosted Higgs boson decays, whilst background processes generally give a more flat distribution.



(a) BM1 Mass Scan.



(b) BM2 Mass Scan.

Figure 7.5: Normalised double-b-tag score distribution comparison between example signal mass points and SM background processes for the BM1 and BM2 mass scans.

Figure 7.6 shows the 2D distribution of D_{bb}^{A} and D_{bb}^{B} for an example signal mass point from the BM1-type mass scan, with $M_{\tilde{\chi}_1^0} \approx 1$ GeV and $M_{\tilde{q}} = 1800$ GeV, with an H_{T} cut of 3500 GeV applied. It is clear that these two AK8 jets generally have double-b-tag scores very close to 1, with the upper right-hand corner showing a strong peak.

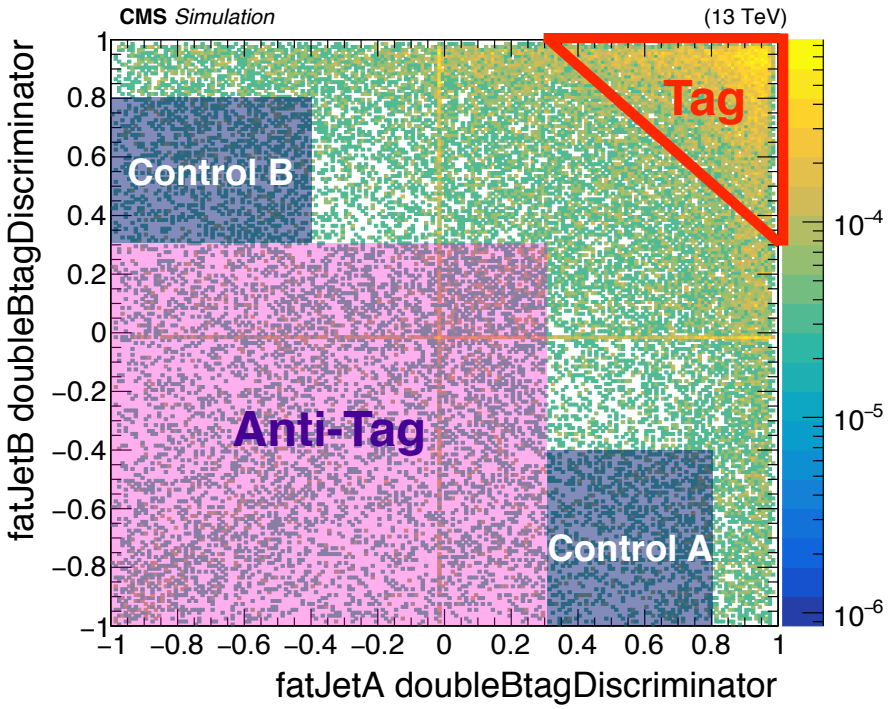


Figure 7.6: Normalised 2D distribution of two highest AK8 jet double-b-tag discriminator scores for an example signal mass point from the BM1 mass scan.

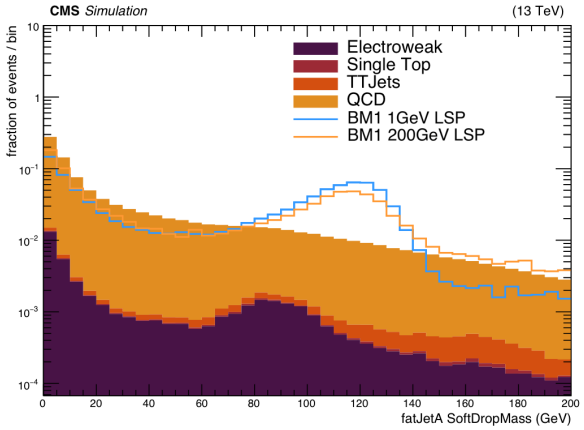
In order to select a large fraction of signal events whilst rejecting as much background as possible, the acceptance region, denoted the "Tag" region, in this 2D plane is defined as the area satisfying equation 7.1.

$$D_{bb}^{B} > 1.3 - D_{bb}^{A} \tag{7.1}$$

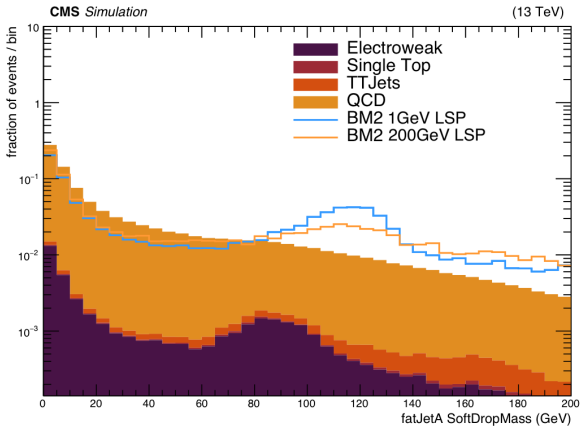
For illustrative purposes figure 7.6 shows the triangular "Tag" region as well as two control regions and an "Anti-tag" region. These remaining regions are used for estimating QCD background contributions as well as verifying other background MC simulation processes, as is detailed in Section 7.5.

7.2.4 2D AK8 soft-drop Mass Binning

Having selected events whose leading two AK8 jet double-b-tag scores reside in the acceptance region defined in Section 7.2.3, we consider the reconstructed soft-drop mass [93] of these AK8 jets, defined in Section 5.5.3. Figure 7.7 shows a



(a) BM1 Mass Scan.



(b) BM2 Mass Scan.

Figure 7.7: Normalised soft-drop mass distribution comparison between example signal mass points and SM background processes for the BM1 and BM2 mass scans.

comparison of soft-drop mass of AK8 jet A between various signal mass points and SM background processes.

We see here a strong peak around 125 GeV, the mass of an SM-like Higgs boson, across the mass points in these NMSSM scenarios. SM background processes such as $t\bar{t}$ and those involving W and Z bosons also show a peak around $80 - 90$ GeV, with QCD giving an exponentially decaying distribution without structure.

We also note a sharp peak for very low values of soft-drop mass. This is since an AK8 jet containing two bottom quarks but where one has been erroneously

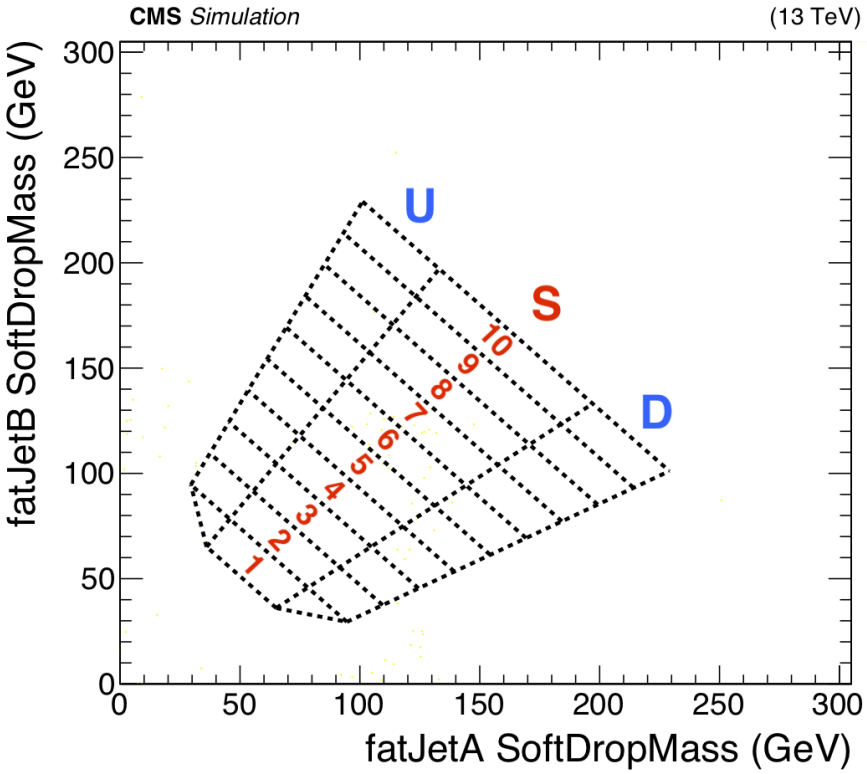


Figure 7.8: Central and sideband binning in the 2D AK8 soft-drop mass plane, as used in this analysis.

removed by the pruning techniques employed by the soft-drop algorithm will have a mass of around 5 GeV, the bottom quark mass.

This analysis contains ten adjacent rectangular signal bins, aligned along the $y = x$ diagonal in the 2D soft-drop mass plane, as shown in Figure 7.8. Each central signal bin S_i is accompanied by an upper and lower sideband bin, labelled U_i and D_i respectively, where $i \in [1, 10]$. These sidebands, situated away from the diagonal, are expected to contain relatively few signal events, since signal events featuring two boosted $H \rightarrow b\bar{b}$ pairs would populate mainly the region around $(125\text{GeV}, 125\text{GeV})$. However QCD background events, given a falling spectrum without any peaks or discernible structure, should populate the sidebands and signal bins in such a way that the ratio of events between the sideband and central region bins would be expected to remain approximately constant as the soft-drop mass increases. Therefore the sidebands are used in order to derive data-driven QCD event yields, as will be detailed in section 7.5.

The original analysis in Ref [2] was designed to search for a light scalar Higgs boson with mass between 30 and 125 GeV. However for the purposes of this work the Higgs boson is that of the Standard Model, with a known mass of 125 GeV which falls at the far end of the original mass range. Therefore in order to better distinguish these NMSSM signal scenarios from the various background processes the soft-drop mass bins, along with their respective sidebands, are modified. As such, the central values of the ten measurement bins used in this analysis, as shown in Figure 7.8, cover a soft-drop mass range from $\sim$ 60 GeV to $\sim$ 160 GeV, in order to fully encapsulate any peaks in the event yield distributions for each of the signal mass points. Given the fairly wide peaks in the soft-drop mass distributions for signal processes, shown in Figure 7.7, the mass bins are accordingly quite wide. It is expected for signal events containing SM-like Higgs bosons to populate predominantly bins 6, 7 and 8 in Figure 7.8.

Figures 7.9 to 7.11 show the 2D AK8 jet soft-drop mass distributions for an example signal mass point in the BM1 mass scan, along with those for QCD and $t\bar{t}$ background processes.

As is shown in figure 7.9 a large fraction of signal events fall into the central signal region bins in the 2D soft-drop mass plane. This example mass point is taken from the BM1 mass scan, with a squark mass of 1800 GeVand LSP mass of 1 GeV. However, the distribution is very similar for other mass points across the six mass scans.

We see in Figure 7.10 QCD events display no real structure across the 2D soft-drop mass plane, with most events residing near either axis or the origin. The remaining three background categories show varying levels of structure, with $t\bar{t}$ MC events containing AK8 jets with soft-drop mass around that of the W and Z bosons, as well as near the top quark mass.

Electroweak and single top quark events show a similar structure, albeit to a lesser extent, mainly populating regions near the horizontal and vertical axes. Furthermore the processes contributing to the electroweak category generally have

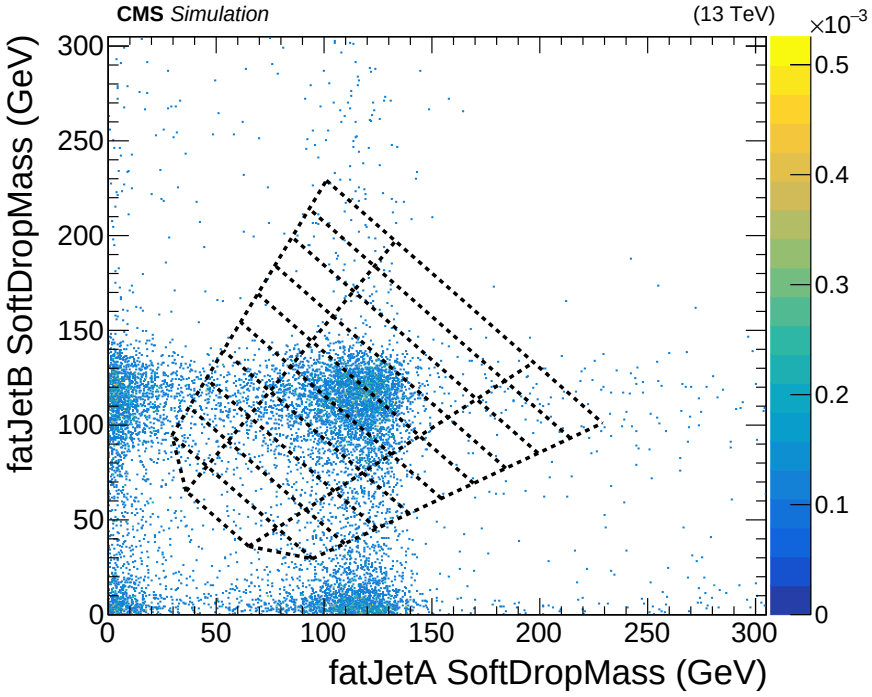


Figure 7.9: Normalised 2D soft-drop mass distribution for the two AK8 jets with highest double-b-tag discriminator scores for an example signal mass point from the BM1 mass scan.

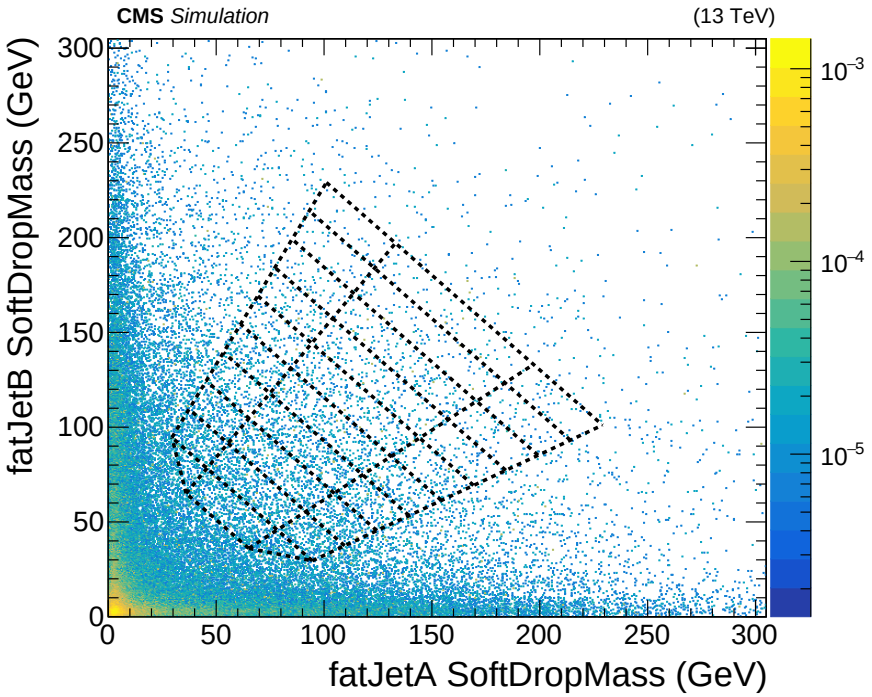


Figure 7.10: Normalised 2D soft-drop mass distribution for the two AK8 jets with highest double-b-tag discriminator scores for QCD MC.

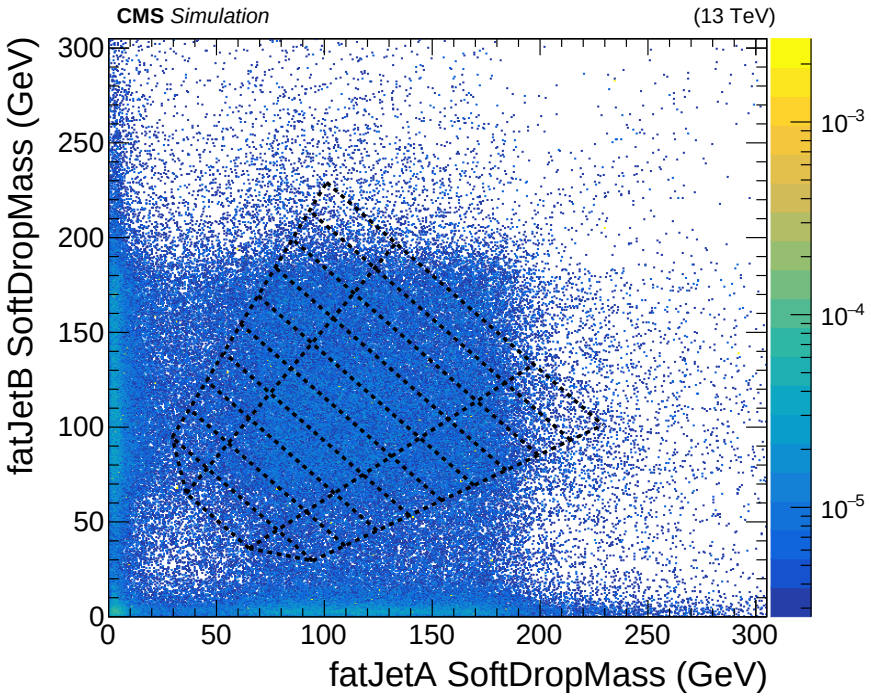


Figure 7.11: Normalised 2D soft-drop mass distribution for the two AK8 jets with highest double-b-tag discriminator scores for $t\bar{t}$ MC.

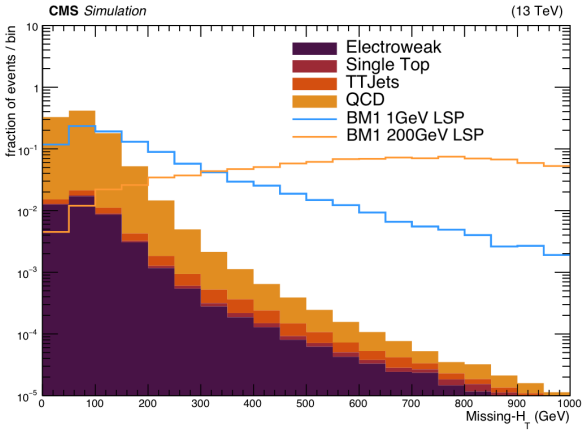
much lower cross-sections than QCD and $t\bar{t}$, and so are not expected to give significant contributions to overall event yields.

7.2.5 Extended Event Selection

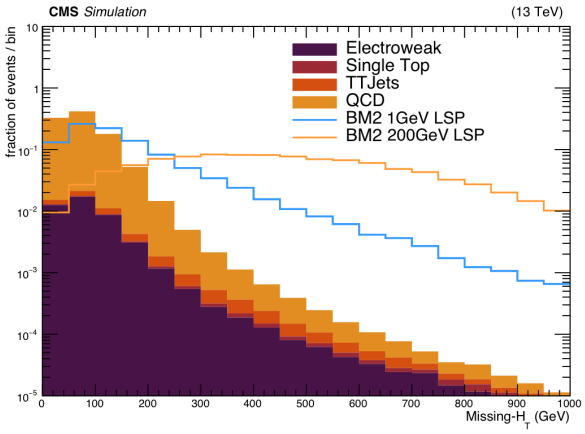
The comparisons shown in Figures 7.2 to 7.5 and 7.7 demonstrate that the event selection criteria in this analysis are appropriate for distinguishing between contributions from SM background processes and the NMSSM scenarios under investigation in this thesis, given the respective distributions peak at different values. However, in order to increase sensitivity as the LSP mass increases towards 200 GeV we introduce additional event selection criteria, in addition to those detailed thus far from Ref [2], detailed in Sections 7.2.6 and 7.2.7.

7.2.6 $H_{\mathrm{T}}^{\mathrm{miss}}$ Binning

Whilst a very light LSP will highly suppress any observed $H_{\mathrm{T}}^{\mathrm{miss}}$, increasing this mass quickly generates a larger $H_{\mathrm{T}}^{\mathrm{miss}}$ signal, especially in the case where the LSP



(a) BM1 Mass Scan.



(b) BM2 Mass Scan.

Figure 7.12: Normalised $H_{\mathrm{T}}^{\mathrm{miss}}$ distribution comparison between example signal mass points and SM background processes for the BM1 and BM2 mass scans.

is heavier than the Higgs boson, thus inheriting the lion's share of the NLSP momentum. This is illustrated in Figure 7.12.

Therefore, in order to better probe the breadth of parameter space for which $0 < M_{\mathrm{LSP}} < 200$ GeV, we split the $H_{\mathrm{T}}^{\mathrm{miss}}$ range into two bins:

$$H_{\mathrm{T}}^{\mathrm{miss}} \in [0, 200), [200, \infty) \text{ GeV}.$$

It may be seen in Figure 7.12 that for mass points with very light LSP the majority of events will fall into the lower $H_{\mathrm{T}}^{\mathrm{miss}}$ bin. However, since this will also be the case for the main background processes we expect a smaller number of background

events in bins for which $H_{\mathrm{T}}^{\mathrm{miss}} \in [200, \infty)$ GeV. As a result, sensitivity to scenarios with heavier LSP masses should benefit considerably.

7.2.7 Veto on Isolated Leptons, Photons and Tracks

In order to reduce contributions from events featuring leptons, especially in the $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV region where the $t\bar{t}$ yield increases relative to that of other background processes, a veto is applied on events containing isolated leptons, isolated photons and isolated tracks.

These objects, defined along with the relevant isolation criteria in Section 5.5, do not generally appear in signal MC events where both Higgs bosons decay into bottom quark-antiquark pairs. Furthermore, signal events containing Higgs bosons decaying leptonically would not be expected to reside within the acceptance regions of the 2D AK8 jet double-b-tag score plane, and as such are not expected to pass the full kinematic event selection criteria.

However background processes such as $t\bar{t}$ and vector boson production may feature a number of isolated leptons originating from particle decays. Furthermore misreconstructed, non-prompt leptons may appear as isolated tracks in data. Therefore removing events containing isolated leptons, photons and tracks helps suppress background contributions, especially in measurement bins featuring high $H_{\mathrm{T}}^{\mathrm{miss}}$– for example reducing the number of electroweak events by more than 50% – whilst having negligible effect on overall signal yields.

7.3 Event Triggers

Given the lowest H_{T} requirement across the entire event selection criteria in this analysis is 1500 GeV, triggering upon H_{T} alone with a threshold a little below this allows for high efficiency, reducing background contributions considerably whilst retaining essentially all signal events due to the respective H_{T} distributions shown in Section 7.2.

Figure 7.13 shows the triggering efficiency as a function of the offline H_{T} for the SingleMuon dataset of run years 2016 and 2017. This is defined as the fraction of events whose HLT H_{T} passes a run year-dependent threshold: $H_{\mathrm{T}}^{\mathrm{HLT}} > 900$ GeV for run year 2016 and $H_{\mathrm{T}}^{\mathrm{HLT}} > 1050$ GeV for run year 2017.

We see that despite the different thresholds for run years 2016 and 2017, for offline H_{T} exceeding 1500 GeV, as is used in this analysis, the efficiency of the H_{T} trigger for both years is 100 % efficient.

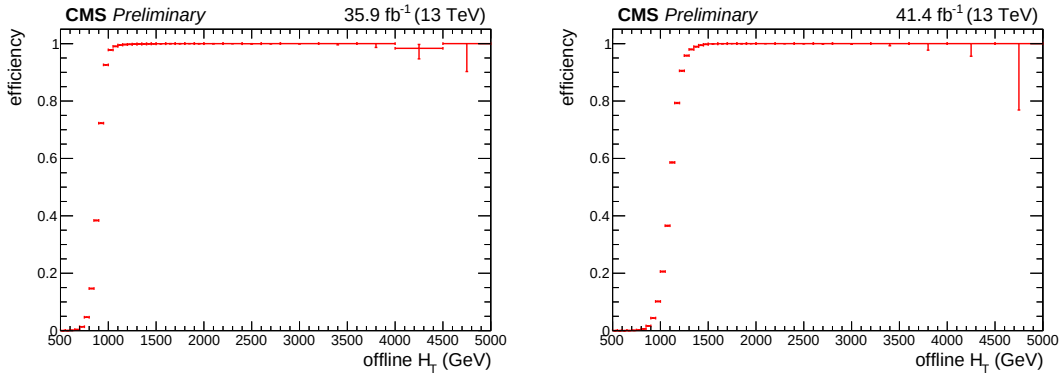
7.3.1 Triggers for 2016 Run Year

Initially a simple H_{T} trigger alone was to be employed, with a threshold of 900 GeV since this should have essentially 100 % efficiency after applying an offline H_{T} cut of 1500 GeV as shown in Figure 7.13. However owing to an H_{T} triggering inefficiency issue in run period H of 2016, some events during this period with H_{T} greater than 1500 GeV may cause saturated trigger towers, thus escaping this trigger.

In order to mitigate this inefficiency therefore, a logical inclusive OR is taken between the following triggers:

- `HLT_PFHT900`,
- `HLT_AK8_PFJet450`,

with these trigger thresholds chosen since these are the lowest thresholds where no prescales were applied for 2016 data taking, and to ensure 100 % efficiency for events with offline $H_{\mathrm{T}} > 1500$ GeV.



(a) H_{T} trigger efficiency for run year 2016. (b) H_{T} trigger efficiency for run year 2017.

Figure 7.13: H_{T} trigger efficiency plots for the SingleMuon dataset of run years 2016 and 2017, with HLT H_{T} trigger thresholds 900 and 1050 GeV respectively.

7.3.2 Triggers for 2017 Run Year

The triggers used for the 2017 run year are essentially the same as those used for 2016; an H_{T} trigger and a single jet p_{T} trigger. However in this case the respective thresholds have been raised to 1050 GeV and 500 GeV, in accordance with the increased instantaneous luminosity at CMS from the LHC. Thus the logical OR of the following triggers is taken:

- `HLT_PFHT1050`,
- `HLT_AK8_PFJet500`,

As shown in Figure 7.13, H_{T} trigger efficiency of 100 % is reached at a higher offline H_{T} value compared with 2016, due to the higher threshold for run year 2017, however in both cases triggering is 100 % efficient for offline $H_{\mathrm{T}} > 1500$ GeV.

7.3.3 Level-1 Prefiring Problem

Across both the 2016 and 2017 run years, the CMS Level-1 trigger suffered from a minor issue known as prefiring, the root of which being a gradual timing shift in ECAL which was not propagated properly to the Level-1 trigger primitives. This issue was found to have minimal effect on the event yields in this analysis, with uncertainties derived in Section 7.7.4 following the approved recipe [104].

7.4 Data and MC Simulated Processes

7.4.1 Data

This analysis uses data collected across run years 2016 and 2017 at the CMS detector, corresponding to integrated luminosities of 35.92 ± 0.90 fb^{-1} and 41.53 ± 0.96 fb^{-1} respectively. These uncertainties correspond to relative values of 2.5 % and 2.3 % respectively, derived from analysis of Van der Meer scans performed during each run year [105, 106].

In both years the datasets used for the primary analysis is JetHT, with the SingleMuon dataset used for the trigger studies in section 7.3. The following JSON files are used to identify runs containing the data which has been certified as good for use in analyses:

- 2016: `Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt`,
- 2017: `Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt`.

Each of run years 2016 and 2017 is split into eras, with the 2016 dataset comprising eras B–H and 2017 B–F, with era A of a given run year usually containing test and calibration runs, and no physics data.

7.4.2 Background Composition

As is the case in Ref [2], background processes are expected to be the production of QCD multijet events, followed by top quark-antiquark production along with extra hadronic jets. Additional sub-dominant yet non-negligible background processes are W-boson and Z-boson production, each featuring extra hadronic jets.

In addition to these processes we also take into the production of a single (anti-)top quark (ST), as well as $t\bar{t}W$, $t\bar{t}Z$ and di-boson (WW, WZ and ZZ) production. The W+jets, Z+jets, $t\bar{t}W$, $t\bar{t}Z$ and di-boson processes are combined, denoted as electroweak (EWK) background. However, it is unlikely that any of these will

remain a significant source of background events following the kinematic event selection criteria, owing especially to the requirement of two double-b-tagged AK8 jets and the respective soft-drop mass cuts. AK8 jets containing only one bottom quark which are accidentally double-b-tagged would be expected to have a low soft-drop mass, around the bottom quark mass, and so would lie close to the axis in the 2D soft-drop mass plane, outside the signal mass regions.

7.4.3 Background MC Samples

Table 7.1 shows each of the background MC processes considered in this analysis, along with their production cross-section values. These samples were centrally produced using CMSSW releases 8_0_X and 9_4_X for 2016 and 2017 respectively, with the double-b-tag scores updated using a newer version of the tagger than is present in the miniAOD files [103]. The production cross-section values shown are taken from the CMS cross-section database (XSDB), with the exception of $t\bar{t}$ whose cross-section is calculated at NNLO+NNLL by the LHCXSWG [107].

Table 7.1: Standard Model background processes considered in this analysis, in decreasing order of production cross-section.

	Process	$H_{\mathrm{T}}^{\mathrm{LHE}}$ range	σ (pb)
QCD		$1000 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 1500$ GeV	1.206×10^{3}
		$1500 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 2000$ GeV	1.204×10^{2}
		$H_{\mathrm{T}}^{\mathrm{LHE}} \geq 2000$ GeV	2.525×10^{1}
$t\bar{t}$		Inclusive	8.318×10^{2}
		$600 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 800$ GeV	2.667×10^{0}
		$800 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 1200$ GeV	1.098×10^{0}
		$1200 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 2500$ GeV	1.988×10^{-1}
		$H_{\mathrm{T}}^{\mathrm{LHE}} \geq 2500$ GeV	2.368×10^{-3}
ST	$\bar{\mathrm{t}}$: t-channel	Inclusive	8.095×10^{1}
	t: t-channel	Inclusive	1.360×10^{2}
	t: s-channel	Inclusive	1.032×10^{1}
	$\bar{\mathrm{t}}W$	Inclusive	3.085×10^{1}
	tW	Inclusive	3.085×10^{1}
EWK	W+jets ($q\bar{q}'$)	$H_{\mathrm{T}}^{\mathrm{LHE}} \geq 600$ GeV	9.514×10^{1}
	W+jets ($l\nu$)	$800 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 1200$ GeV	6.656×10^{0}
	W+jets ($l\nu$)	$1200 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 2500$ GeV	1.608×10^{0}
	W+jets ($l\nu$)	$H_{\mathrm{T}}^{\mathrm{LHE}} \geq 2500$ GeV	3.891×10^{-2}
	Z+jets ($q\bar{q}$)	$H_{\mathrm{T}}^{\mathrm{LHE}} \geq 600$ GeV	5.279×10^{1}
	Z+jets (ll)	$800 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 1200$ GeV	7.754×10^{-1}
	Z+jets (ll)	$1200 \leq H_{\mathrm{T}}^{\mathrm{LHE}} < 2500$ GeV	1.862×10^{-1}
	Z+jets (ll)	$H_{\mathrm{T}}^{\mathrm{LHE}} \geq 2500$ GeV	4.385×10^{-3}
	WW	Inclusive	6.430×10^{1}
	WZ	Inclusive	2.343×10^{1}
	ZZ	Inclusive	1.016×10^{1}
	$t\bar{t}Z$+jets ($q\bar{q}$)	Inclusive	5.297×10^{-1}
	$t\bar{t}Z$+jets ($ll/\nu\nu$)	Inclusive	2.529×10^{-1}
	$t\bar{t}W$+jets ($q\bar{q}'$)	Inclusive	4.050×10^{-1}
	$t\bar{t}W$+jets ($l\nu$)	Inclusive	2.043×10^{-1}

7.5 Data-Driven Estimation of Background from QCD Multijet Events

Since in this analysis the final state under consideration is all-hadronic, and as the QCD production cross-section relative to other background processes is very large, QCD multijet processes constitute the dominant background. Due to the enormous complexity in simulating QCD events to arbitrarily high accuracy, a data-driven approach was derived to provide a more robust estimate of the QCD yields across the signal region binning [2].

7.5.1 Tag, Anti-Tag and Control Region Binning

In this extension of the analysis in Ref [2] the signal region is divided into 60 measurement bins: two $H_{\mathrm{T}}^{\mathrm{miss}}$ regions each containing three H_{T} bins, each of which is made up of ten 2D AK8 soft-drop mass regions. Accordingly, in each of the double-b-tag regions defined in Figure 7.6 we have central mass bins S_i^{region} and upper and lower sideband bins U_i^{region} and D_i^{region} respectively.

We therefore denote the signal measurement bins and respective upper and lower sideband bins for each double-b-tag region as follows:

- Tag double-b-tag region: S_i^{tag}, U_i^{tag} and D_i^{tag},
- Anti-Tag double-b-tag region: S_i^{anti}, U_i^{anti} and D_i^{anti},
- Control double-b-tag region: S_i^{control}, U_i^{control} and D_i^{control},

where the areas of each S_i equal the sum of the areas of the respective U_i and D_i bins, with the exception of the lowest mass region in which the U_i and D_i bins sum to half the area of S_i. The observed event yield in a given bin is represented by a hatted symbol:

- Tag double-b-tag region: $\hat{S}_i^{\mathrm{tag}}$, $\hat{U}_i^{\mathrm{tag}}$ and $\hat{D}_i^{\mathrm{tag}}$,

- Anti-Tag double-b-tag region: $\hat{S}_i^{\text{anti}}$, $\hat{U}_i^{\text{anti}}$ and $\hat{D}_i^{\text{anti}}$,
- Control double-b-tag region: $\hat{S}_i^{\text{control}}$, $\hat{U}_i^{\text{control}}$ and $\hat{D}_i^{\text{control}}$.

7.5.2 Derivation of QCD Expected Yields

In order to derive data-driven estimations for the QCD yields across the S_i^{tag} bins, predicted yields, denoted $\hat{S}_i^{\text{tag PRED}}$, are calculated using Equation (7.2).

$$\hat{S}_i^{\text{tag PRED}} = (\hat{U}_i^{\text{tag}} + \hat{D}_i^{\text{tag}}) \cdot F_i, \tag{7.2}$$

where

$$F_i := \frac{\hat{S}_i^{\text{anti}}}{\hat{U}_i^{\text{anti}} + \hat{D}_i^{\text{anti}}}, \tag{7.3}$$

with the quantities in Equation (7.3) referring to the anti-tag region illustrated in Figure 7.6. This is possible since the $\hat{S}_i^{\text{anti}}$, $\hat{U}_i^{\text{anti}}$ and $\hat{D}_i^{\text{anti}}$ event yields are completely dominated by QCD events, with negligible contributions from other background or signal processes.

In the perfect case where the density of QCD events in each signal bin S_i^{tag} equals that in its respective sidebands U_i^{tag} and D_i^{tag}, the central bin yields would equal that of the two sidebands summed, i.e. $F_i = 1$; with the exception of the lowest mass region for which the sideband bins are half the area of the central bin, therefore containing half as many events.

However, this is not the case. The relative density of background events increases sharply for mass values very close to the origin, as shown in Figure 7.7, as AK8 jets containing one or no bottom quarks, accidentally double-b-tagged, will have a very low soft-drop mass. Therefore since the central region lies further from either axis than the sideband regions, the ratio between the numbers of events in the sideband and central regions increases as we approach the origin. In order to avoid any effect resulting from this the sideband regions of the first soft-drop mass bin have a triangular shape, and so the F_i factors are required to account for this.

When calculating F_i, the yields in the *anti-tag* double-b-tag region may be used without compensating for events originating from signal or other background processes, due to the complete QCD dominance across all bins in this region. However, the event yields in the *tag* double-b-tag region are much more sensitive to such contamination. As such the $\hat{U}_i^{\text{tag}}$ and $\hat{D}_i^{\text{tag}}$ yields correspond to the observed yields in these bins with expected signal and other background yields, from $t\bar{t}$, single top and electroweak processes, from MC simulated events, subtracted.

7.5.3 Deriving a Robust QCD Estimation

The F_i correction factors derivation method in Section 7.5.2 is well motivated in the case where a very large number of events remain after the kinematic event selection is imposed. However, since the data and even some MC samples suffer from poor statistics in the high-H_{T}, high-$H_{\text{T}}^{\text{miss}}$ regions, despite the QCD and $t\bar{t}$ samples being generated such that they each contain in excess of $1,000,000$ events with $H_{\text{T}} > 2500$ GeV, the statistical uncertainty begins to dominate.

As such, a more stable method was derived in order to reduce the impact of such low statistics, by approximating the expected background yields using a functional form. Since the two AK8 jets leading in double-b-tag score, denoted A and B, are randomised, the event densities in the AK8 jet soft-drop mass plane are invariant under swapping these jets: A$\leftrightarrow$B. This then allows for the soft-drop mass distribution of one of the AK8 jets to be used to describe the two-dimensional soft-drop mass distribution of both AK8 jets A and B.

Therefore rather than calculating ratios of event yields calculated from 2D acceptance regions, we instead calculate the integral of ρ, the event density across the 2D soft-drop mass plane.

Using this method, equation 7.3 becomes the following:

$$F_i = \frac{\iint_{S_n} \mathrm{d}m_a \mathrm{d}m_b \cdot \rho_h^{\text{anti-tag}}(m_a, m_b)}{\iint_{U_n+D_n} \mathrm{d}m'_a \mathrm{d}m'_b \cdot \rho_h^{\text{anti-tag}}(m'_a, m'_b)}, \tag{7.4}$$

where $i = 10((h-1)+(m-1))+n$, $h \in \{1,2,3\}$ is the index of the H_{T} measurement bin, $m \in \{1,2\}$ is the index of the $H_{\mathrm{T}}^{\mathrm{miss}}$ measurement bin and $n \in [1,10]$ the index of the mass region in the 2D soft-drop mass plane.

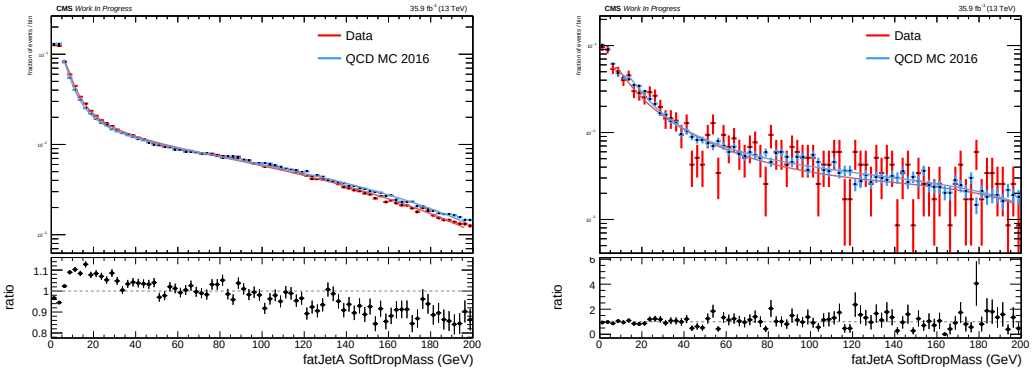
This event density ρ is calculated as the 2D product of the form $\rho(m_1, m_2) = f(m_1)f(m_2)$. Here $f(m)$ is a function fitted to the soft-drop mass distribution of one AK8 jet in a given H_{T}-$H_{\mathrm{T}}^{\mathrm{miss}}$ region, using ROOT and minimising a chi-squared distribution, with all fits having $\chi^2 < 0.1$ and most exhibiting $\chi^2 \ll 0.05$. The polynomial function $f(m)$ is defined as follows, with the number of terms chosen to allow enough freedom in the fit whilst not over-fitting:

$$\begin{aligned} f(m) = \; & p_0 + \frac{p_1}{(m-p_2)} + \frac{p_3}{(m-p_4)^2} + \frac{p_5}{(m-p_6)^3} + \frac{p_7}{(m-p_8)^4} \\ & + p_9(m-p_{10}) + p_{11}(m-p_{12})^2. \end{aligned} \tag{7.5}$$

Since the anti-tag double-b-tag region is strongly QCD-dominated, we may derive these $f(m)$ fit functions using data. However, in both $H_{\mathrm{T}}^{\mathrm{miss}}$ regions within the highest H_{T} region, as well as the $2500 < H_{\mathrm{T}} < 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV region, there are too few events remaining. As such in these regions MC simulated QCD events are used instead, since the anti-tag region is completely dominated by QCD multijet events, with good agreement shown up to the limit of event statistics. Figure 7.14 shows example fits for two of the six H_{T}-$H_{\mathrm{T}}^{\mathrm{miss}}$ regions, with the remainder shown in appendix B.1. In Section 7.5.4 we test this QCD estimation method, and motivate an estimation for the uncertainty on the resultant expected QCD event yields.

7.5.4 Testing Improved QCD Estimation Method

The F_i factors, as calculated in Equation (7.4), are shown in Figure 7.15 for the 60 signal region bins. Figures 7.15a and 7.15b show the F_i factors for run year 2016 calculated using this integration method, compared with those derived simply from the event yields in the 2D central and sideband regions of the anti-tag double-b-tag



(a) $1500 < H_{\mathrm{T}} < 2500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2016.

(b) $H_{\mathrm{T}} > 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2016.

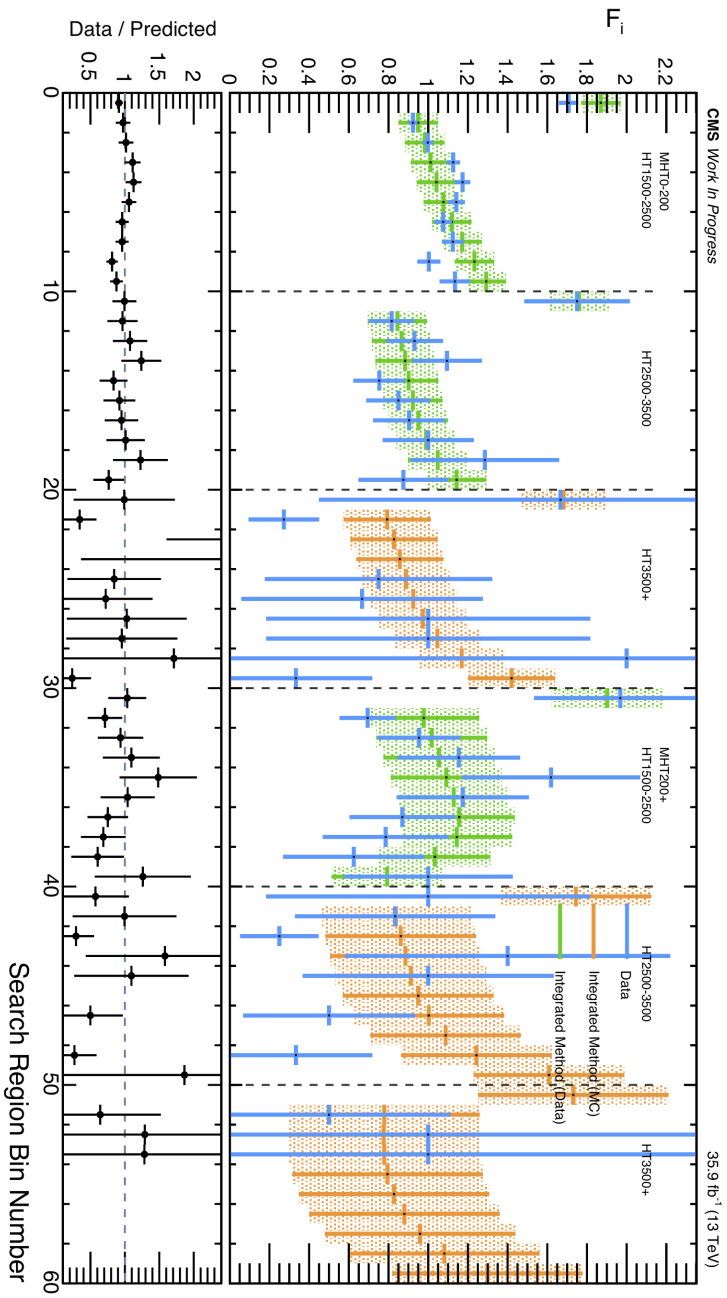
Figure 7.14: Example $f(m)$ fits across the one-dimensional soft-drop mass distributions for data and MC events in the anti-tag region for run year 2016.

score region using data and MC events respectively. Figures 7.15c and 7.15d show the same comparisons, but for run year 2017.

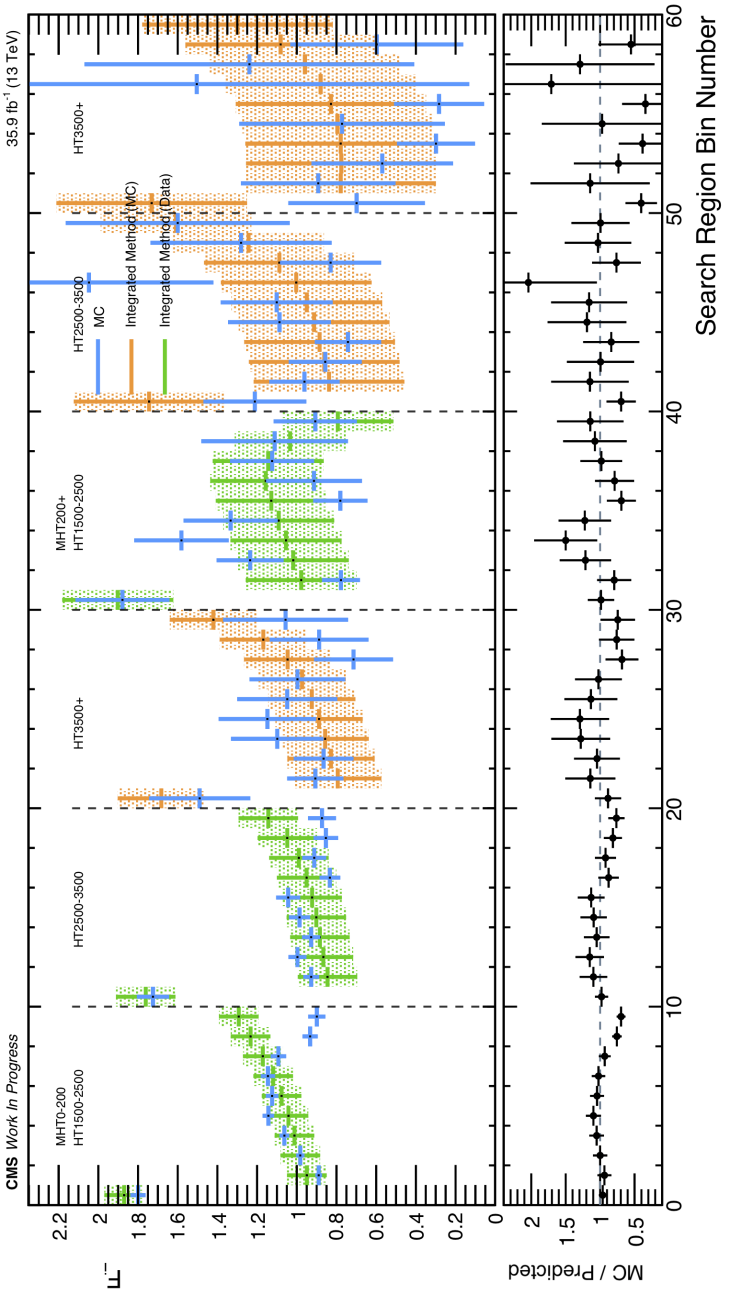
Here the green shaded bands represent F_i factors calculated using Equation (7.4) with the function $f(m)$ fitted to the soft-drop mass observed in data for the respective H_{T}-$H_{\mathrm{T}}^{\mathrm{miss}}$ region. The orange bands represent F_i factors from the same method, but using MC events due to low statistics in data.

We note that the F_i factors tend to be close to unity, with the exception of the first mass bin in each H_{T}-$H_{\mathrm{T}}^{\mathrm{miss}}$ region, which is closer to two. The reason for this lies in the shape of the 2D mass sidebands; the first mass region has sideband bins half the size relative to remaining sideband bins.

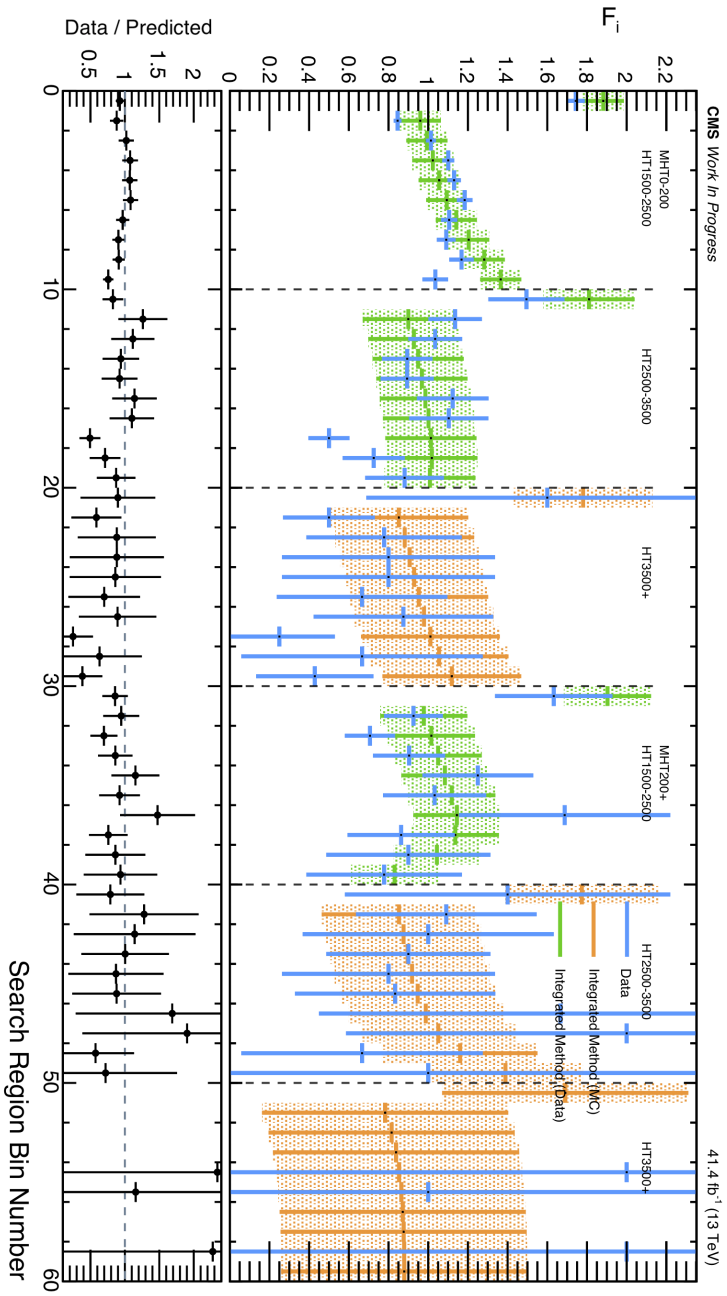
In general we see good agreement across the entire binning, with low data and MC statistics affecting the high-H_{T} and high-$H_{\mathrm{T}}^{\mathrm{miss}}$ regions in particular for the method detailed in equation 7.3. The reason for these low statistics is that following the baseline AK8 jet requirements and high H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$ requirements, only a small fraction of the events remain. These events are then spread out across the 2D soft-drop mass plane and are concentrated towards the origin, meaning very few events fall within the central and sideband regions used in this analysis. However using the integration method shown in Equation (7.4) allows for the soft-drop mass distribution of only one of these AK8 jets to be used, effectively projecting


CMS Work In Progress
35.9 fb-1 (13 TeV)
F_i
Data / Predicted
Search Region Bin Number
MHT0-200
HT1500-2500
HT2500-3500
HT3500+
MHT200+
HT1500-2500
HT2500-3500
HT3500+
Data
Integrated Method (MC)
Integrated Method (Data)

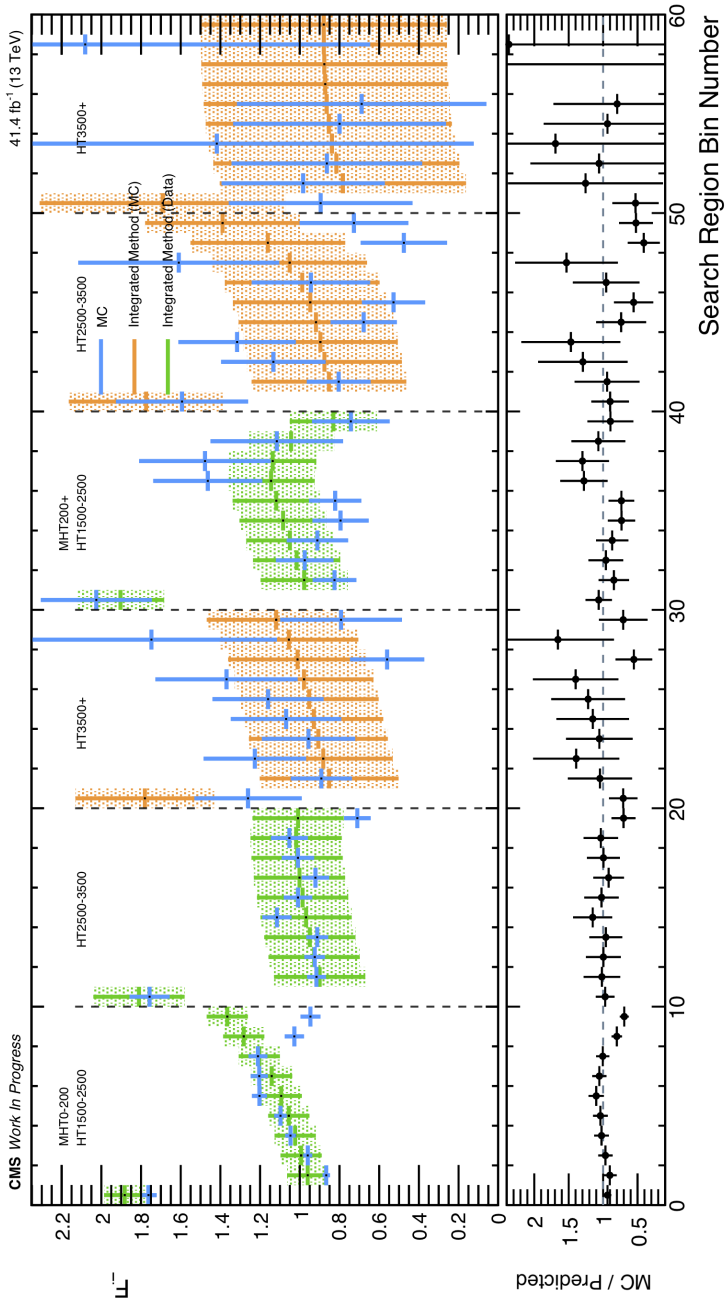
(a) 2016 Data.


CMS Work In Progress
35.9 fb-1 (13 TeV)
MHT0-200
HT1500-2500
HT2500-3500
HT3500+
MHT200+
HT1500-2500
HT3500+
HT2500-3500
MC
Integrated Method (MC)
Integrated Method (Data)
HT3500+
R
MC / Predicted
Search Region Bin Number

(b) 2016 MC.


CMS Work In Progress
41.4 fb^{-1} (13 TeV)
F_i
Data / Predicted
Search Region Bin Number
MHT0-200
HT1500-2500
HT2500-3500
HT3500+
MHT200+
HT1500-2500
HT2500-3500
HT3500+
Data
Integrated Method (MC)
Integrated Method (Data)

(c) 2017 Data.



(d) 2017 MC.

Figure 7.15: QCD F_i calculated values across the sixty signal region bins in the anti-tag DBT region compared with data and QCD MC central/sideband event yield ratios.

all events across the entire 2D soft-drop mass plane onto one axis, resulting in much higher event statistics.

Having such low statistics in certain regions further validates the use of the integration method in Equation (7.4) for deriving the F_i factors used in estimating the QCD event yields, with the F_i factors derived from the 2D regions showing much larger fluctuations due to low event statistics.

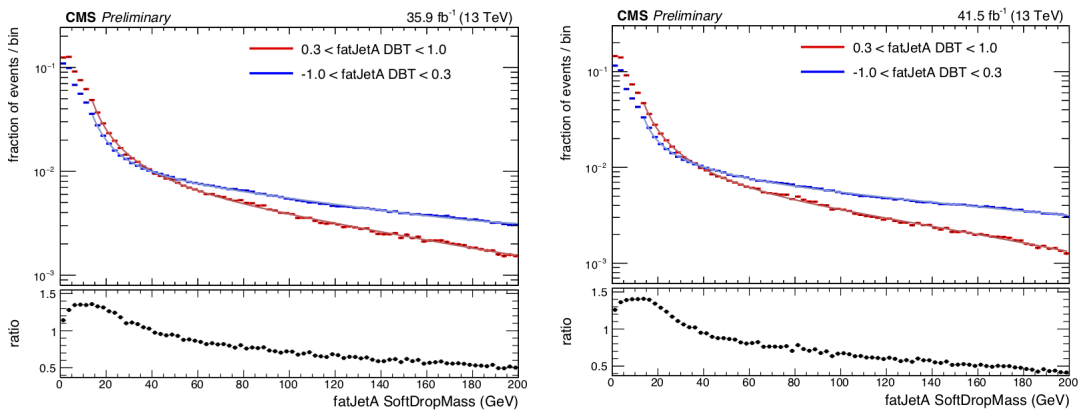
In order to estimate the systematic uncertainty on these F_i factors derived using the integration method, we consider the ratios between these factors and the respective values calculated using the actual event yields in the 2D regions. For each H_{T}-$H_{\mathrm{T}}^{\mathrm{miss}}$ region the standard deviation is calculated across these ratios, with this taken as the value for the $\pm 1\sigma$ uncertainty on the F_i within the given region. We note that for the most part, this standard deviation is very close to the value obtained by treating each of the ratio distributions as a Gaussian and taking half of the width of the 68th percentile about the mean.

These uncertainty values are shown in Figure 7.15 as the error bars of the orange and green bands, and are the propagated through into the datacards when performing the final fit in order to calculate expected and observed limits.

7.5.5 Double-*b*-tag Score Dependence on soft-drop Mass

The method detailed in section 7.5.2 holds under the assumption of no correlation between the soft-drop mass of an AK8 jet and its double-b-tag score: meaning ratios between event yields in the anti-tag region are approximately equal to those in the tag region.

Correction factors are derived, representing the ratio between the F_i factor for a given mass bin in the anti-tag region and the same factor in the tag double-b-tag region. Therefore should these values differ significantly from unity, this method for using the ratio calculated in the anti-tag region to predict event yields in the tag region would not prove valid.



(a) $1500 < H_{\mathrm{T}} < 2500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2016.

(b) $H_{\mathrm{T}} > 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2016.

Figure 7.16: Example $f(m)$ fits across the one-dimensional soft-drop mass distributions for data and MC events in the anti-tag region for run year 2016.

However, we cannot derive these correction factors from data in the tag double-b-tag region without prematurely unblinding the analysis. Therefore in order to assess the extent to which no correlation exists, values for data-driven correction factors were derived using a QCD-dominated control region defined as follows.

- $1500 < H_{\mathrm{T}} < 2500$ GeV.
- Two AK8 jets leading in double-b-tag score must have $p_{\mathrm{T}} > 300$ GeV.
- At least one of the two AK8 jets must have double-b-tag score less than 0.3.

Firstly, using the same approach as in Equation (7.5), a function $f(m)$ is fitted to the 1D distribution of the soft-drop mass of AK8 jet A, in each of the following cases: where the double-b-tag discriminator score for AK8 jet A is less than 0.3 and greater than 0.3, as is shown in Figure 7.16. A 2D product is then calculated as before: $\rho(m_1, m_2) = f(m_1) \cdot f(m_2)$.

The correction factors are denoted C_i, where $i \in \{1, ..., 10\}$, for the ten AK8 soft-drop mass bins in the above control region, and are calculated in Equation (7.6). Only ten bins exist here, since the control region contains only one H_{T} bin and no $H_{\mathrm{T}}^{\mathrm{miss}}$ binning.

$$C_i = \frac{\iint\limits_{S_i} \mathrm{d}m_a \mathrm{d}m_b \cdot \rho_h^{\mathrm{tag}}(m_a, m_b)}{\iint\limits_{U_i+D_i} \mathrm{d}m'_a \mathrm{d}m'_b \cdot \rho_h^{\mathrm{tag}}(m'_a, m'_b)} \Bigg/ \frac{\iint\limits_{S_i} \mathrm{d}m''_a \mathrm{d}m''_b \cdot \rho_h^{\mathrm{anti}}(m''_a, m''_b)}{\iint\limits_{U_i+D_i} \mathrm{d}m'''_a \mathrm{d}m'''_b \cdot \rho_h^{\mathrm{anti}}(m'''_a, m'''_b)}. \tag{7.6}$$

Following Equation (7.6) the C_i estimated values are shown in table 7.2. These values all lie within 10 % of unity, thus validating the use of the ratio of event yields in the anti-tag double-b-tag region in order to estimate yields in the tag region. This estimation of 10% is then added as a systematic uncertainty on the expected QCD event yield when performing the fits in order to calculate the expected and observed limits.

Table 7.2: Estimated values of C_i for run years 2016 and 2017, the correction factors accounting for soft-drop mass dependence in the QCD estimation method.

	2016	2017
C_1	1.016	0.967
C_2	0.944	0.889
C_3	0.935	0.914
C_4	0.936	0.936
C_5	0.941	0.952
C_6	0.947	0.965
C_7	0.956	0.976
C_8	0.967	0.988
C_9	0.982	1.002
C_{10}	1.003	1.017

7.5.6 Testing QCD Estimation in Control Region

In order to evaluate the performance of the QCD estimation method against data, we use a control region defined as the union of the two shaded control regions in the double-b-tag score plane in Figure 7.6. This region stipulates that one of the two AK8 jets must have discriminator score less than 0.4, with the other having a score between 0.3 and 0.8, such that contributions from signal processes are negligible.

Defining F_i using the integration method in Equation (7.4), we define $\hat{S}_i^{\mathrm{control,\ pred}}$, the predicted event yield in each of the central mass bins in this control region, as follows:

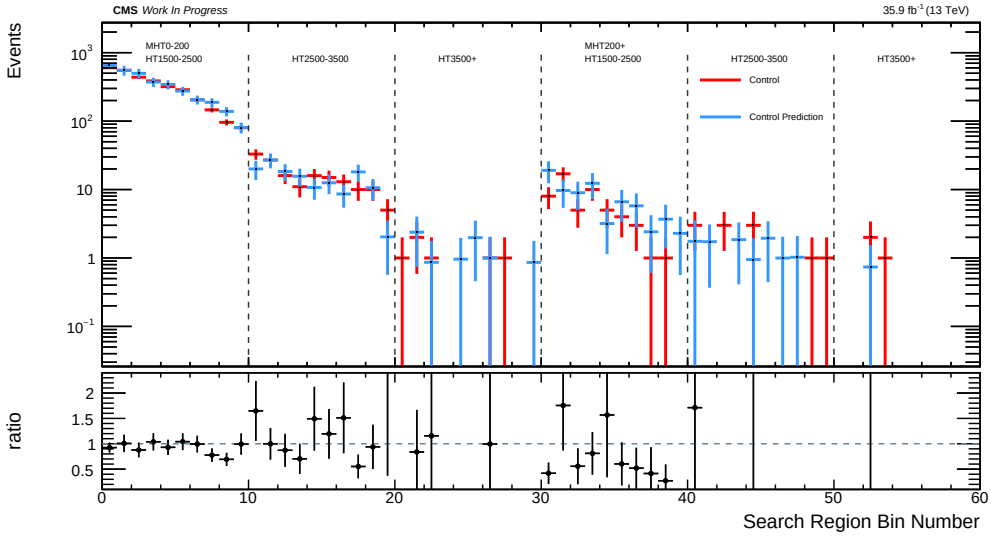
$$\hat{S}_i^{\text{control, pred}} := F_i \cdot (\hat{U}_i^{\text{control}} + \hat{D}_i^{\text{control}}), \tag{7.7}$$

where $\hat{U}_i^{\text{control}}$ and $\hat{D}_i^{\text{control}}$ represent the yields in the mass sidebands. We see in Figure 7.17 a direct comparison between the observed and predicted event yields across the sixty measurement bins in the QCD-dominated control region.

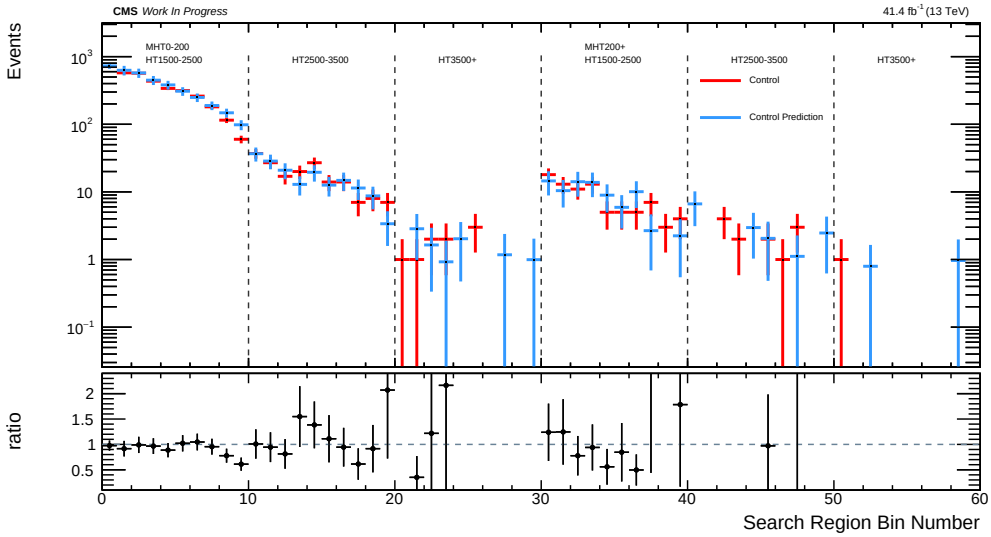
We see in general good agreement between observed and predicted yields, limited somewhat by low statistics in the higher H_{T} and $H_{\text{T}}^{\text{miss}}$ regions.

7.5.7 Conclusions of QCD Studies

Having studied and tested this QCD prediction method, we have found it to be self-consistent, with good agreement shown between F_i factors calculated using both methods, for both data and MC events, in Section 7.5.4. Furthermore, in Section 7.5.6 we found that the event yields predicted using data in a control region agree well with observation, within statistical uncertainty. Given the mass-dependent variations in double-b-tag score, along with these other validation tests, the data-driven QCD event yields have therefore been shown to be reliable to within the uncertainty values applied to the F_i factors and the overall 10 % systematic uncertainty motivated in Section 7.5.5.



(a) Run year 2016.



(b) Run year 2017.

Figure 7.17: Predicted event yield compared with observation in the QCD control region for run years 2016 and 2017 respectively.

7.6 Background from $t\bar{t}$ Production

Background contribution from top quark pair production is estimated using MC generated event samples. These samples have corrections applied which were derived from data, as discussed in this section.

7.6.1 Top Quark p_{T} Reweighting

It has been found that the top quark p_{T} spectrum for MC simulated $t\bar{t}$ events is generally higher than is observed in data due to the finite order to which these events are generated [108]. Therefore in order to account for this a top quark p_{T}-dependent event weighting has been derived for Powheg NLO $t\bar{t}$ MC samples by the CMS top MC working group by parameterising the ratio between observed and MC $t\bar{t}$ events.

This weighting only applies to $t\bar{t}$ MC events, not to signal or other background processes such as $t\bar{t}Z$ or single top quark production. Firstly the p_{T} of each parton level top quark is studied: this is the p_{T} of each top quark after the initial production by the event generator but before it undergoes any treatment at the hands of Pythia, in order to not interfere with the application of any run year-specific calibrations.

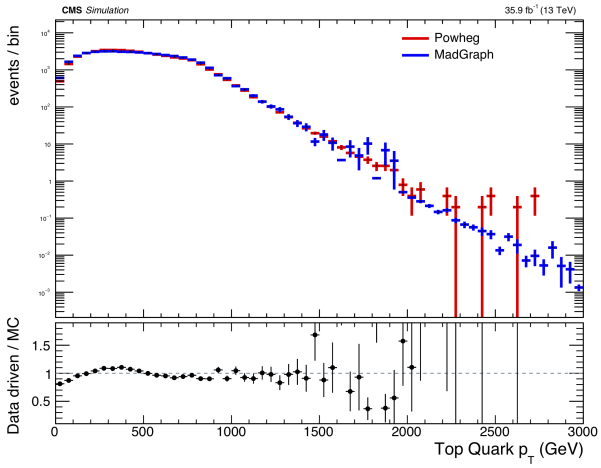
Following the recipe from the CMS top MC working group, a scale factor $S_{p_{\mathrm{T}},i}$ is assigned to each top quark i ($i = 1, 2$) as follows [108]:

$$Sp_{\mathrm{T}}, i = e^{0.0615 - 0.0005(\mathrm{GeV}^{-1}) \cdot p_{\mathrm{T},i}}, \tag{7.8}$$

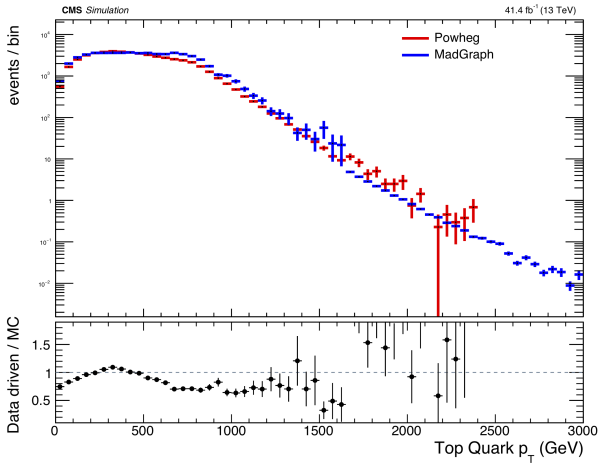
with total event weight w defined as the geometric mean of the two scale factors:

$$w = \sqrt{S_{p_{\mathrm{T}},1} S_{p_{\mathrm{T}},2}}. \tag{7.9}$$

However the purpose of this re-weighting is to correct the shape of the top quark p_{T} distribution, not the overall normalisation. Therefore when calculating event



(a) Powheg and MadGraph 2016.



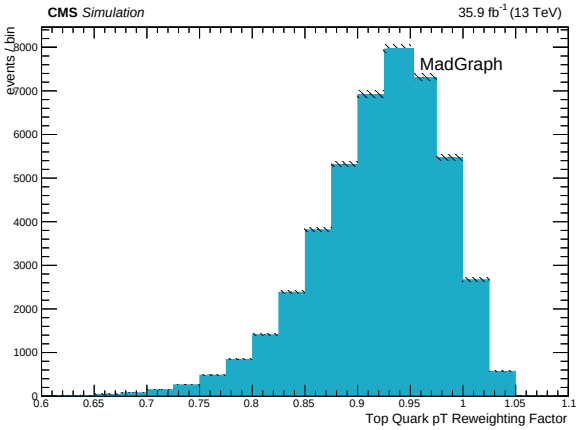
(b) Powheg and MadGraph 2017.

Figure 7.18: Top quark p_{T} distribution comparisons between Powheg and MadGraph MC samples for run years 2016 and 2017.

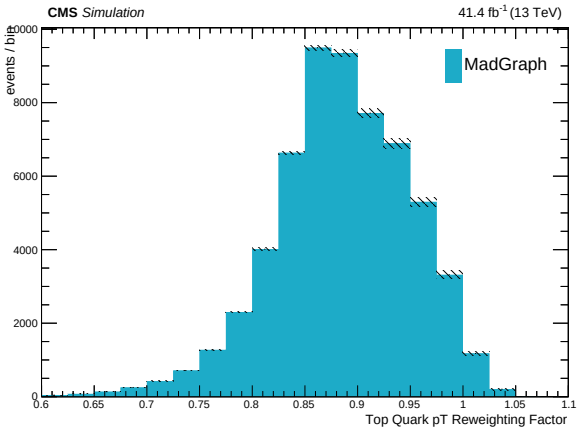
yields the weighting for each event is divided by the average top quark p_{T} weight $\bar{w}$ for the respective MC sample.

The top quark p_{T} distributions for the Powheg and MadGraph $t\bar{t}$ MC samples appear very similar, as is shown in Figure 7.18. As such, the same top quark p_{T} reweighting scale factor calculation in Equation (7.8) is also used for the MadGraph-generated $t\bar{t}$ MC events, which are used in this analysis due to the much larger available equivalent luminosity of the samples. The top quark p_{T} event weighting distributions are shown in Figure 7.19 for run years 2016 and 2017.

.



(a) MadGraph 2016.



(b) MadGraph 2017.

Figure 7.19: Scale factor weights as a function of top quark p_{T} for run years 2016 and 2017.

We note a difference in the top quark p_{T} scale factors between MadGraph LO MC samples for run years 2016 and 2017, as is shown in Figure 7.19. This is due to a slight difference in the configurations used by the CMS MC production working group when generating the MadGraph LO $t\bar{t}$ samples for the respective run years, resulting in slightly different top quark p_{T} distributions.

7.6.2 $t\bar{t}$-enriched Control Region

When using MC simulated events to describe the $t\bar{t}$ background process, it is important to ascertain how well these simulated events represent real $t\bar{t}$ events in data. This is then used in order to estimate the systematic uncertainty on the expected $t\bar{t}$ background yields in this analysis.

Such a control region should ideally satisfy the following criteria:

- Contain a negligible (or otherwise subtractible) amount of events from signal processes.
- Be dominated by the respective background process under examination.
- Contain enough observed events to make for a useful comparison between data and MC.
- Must represent the kinematic properties of the signal region in the analysis.

A control region was constructed in this analysis by defining three measurement bins $\tilde{S}_i$ in the 2D AK8 jet soft-drop mass plane, centering around (m_t, m_t) as shown in Figure 7.20, analogous to the S_i bins in the signal region and complete with respective $\tilde{U}_i$ and $\tilde{D}_i$ sidebands. In this region only the lowest baseline H_{T} cut of 1500 GeV is applied, rather than full H_{T} binning, thus further reducing the signal event yields relative to $t\bar{t}$.

In order to boost the statistics in this control region two adjustments are made. Firstly, bins in this new 2D mass grid have twice the area of those used in the signal

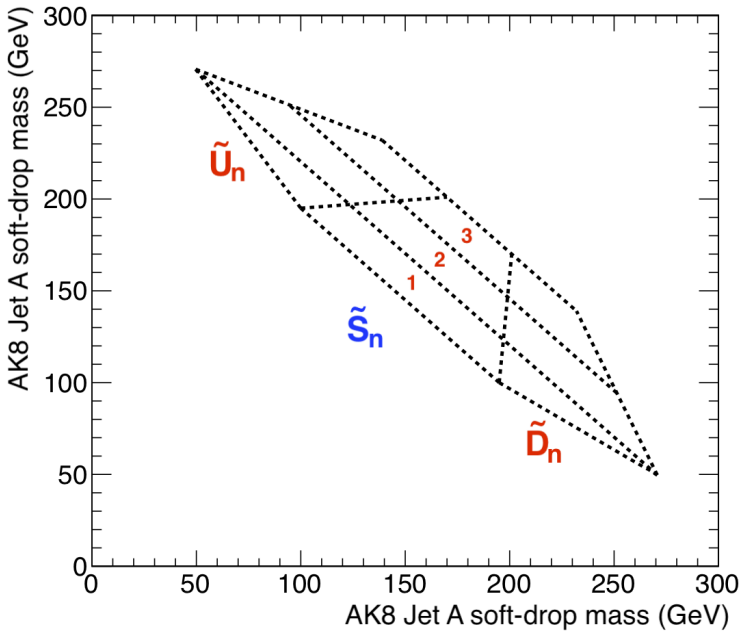


Figure 7.20: Three 2D AK8 soft-drop mass bins defining the $t\bar{t}$ control region.

region of the analysis. Additionally the tag' region is simply that for which the two highest AK8 jet double-b-tag scoring jets score above 0.3, as shown in Figure 7.21. Thus the triangular signal region in the 2D double-b-tag discriminator score plane is replaced by a square of twice the area. This is in order to maximise event statistics for the high H_{T} regions, given the decreasing number of QCD events as we increase the soft-drop masses of the two AK8 jets. We note that the definition of the anti-tag region is unaffected.

We may summarise the event selection requirements for this control region as follows:

- $H_{\mathrm{T}} > 1500$ GeV.
- Both AK8 jets must have double-b-tag discriminator > 0.3.
- AK8 soft-drop mass binning as shown in Figure 7.20.

Substantial contributions from mass points which contain top quarks should not enter this control region, given that higher-momentum Higgs bosons decaying to a bottom quark-antiquark pair should be candidates for the highest double-b-tag

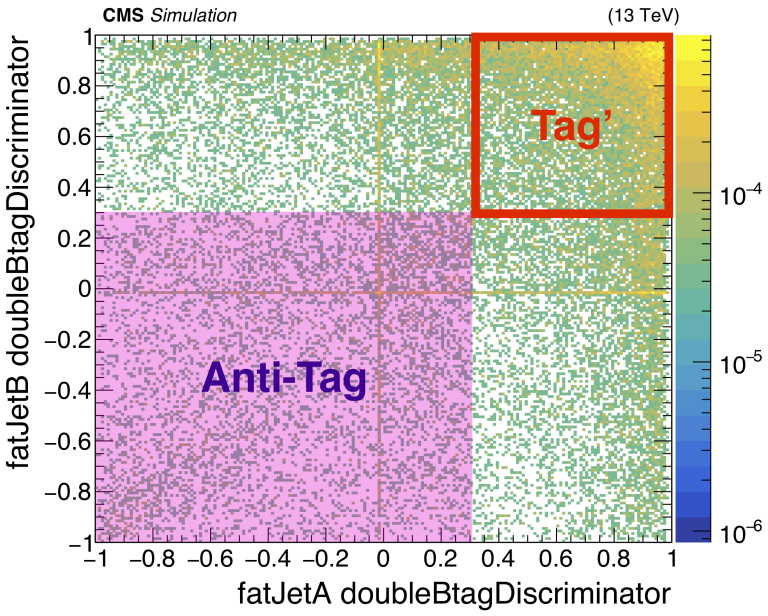


Figure 7.21: Areas in the 2D double-b-tag score plane defining the tag' and anti-tag regions within the $t\bar{t}$ control region.

discriminator output, and the fraction of events containing two top quarks both erroneously double-b-tagged and no $H \to b\bar{b}$ pairs is quite low.

Data-driven estimations are calculated for the $t\bar{t}$ expected yields in this control regions, with the method detailed in Equations 7.10-7.13.

Background events from QCD processes will inevitably feature in this control region, though it is expected that $t\bar{t}$ events will account for a large fraction, as shown in Figure 7.22. As such we calculate a data-driven estimation for the QCD yield which must then be subtracted in order to arrive at an estimate for the $t\bar{t}$ yield.

$$\hat{\tilde{S}}^{\text{ tag'}}_{i,\ t\bar{t}\ \text{estimate}} = \hat{\tilde{S}}^{\text{ tag'}}_{i,\ \text{data}} - \hat{\tilde{S}}^{\text{ tag'}}_{i,\ \text{V, MC}} - \hat{\tilde{S}}^{\text{ tag'}}_{i,\ \text{QCD}} \tag{7.10}$$

where $\hat{\tilde{S}}^{\text{ tag'}}_{i,\ \text{V, MC}}$ represents the yield from MC-simulated W+jets and Z+jets events,

$$\hat{\tilde{S}}^{\text{tag'}}_{i,\ \text{QCD}} = \tilde{F}_i \cdot (\hat{\tilde{U}}^{\text{tag'}}_{i,\ \text{QCD}} + \hat{\tilde{D}}^{\text{tag'}}_{i,\ \text{QCD}}) \tag{7.11}$$

and

$$\tilde{F}_i \equiv \frac{\hat{\hat{S}}^{\text{ anti}}_{i,\text{ QCD}}}{\hat{\hat{U}}^{\text{ anti}}_{i,\text{ QCD}} + \hat{\hat{D}}^{\text{ anti}}_{i,\text{ QCD}}}, \tag{7.12}$$

similarly to as shown for the QCD estimation method in Section 7.5.2.

Writing Equations (7.11) and (7.12) in terms of the observed yields rather than predicted yields, we obtain the following:

$$\hat{\hat{S}}^{\text{tag'}}_{i,\text{ QCD}} = \tilde{F}_i \cdot \Big(\hat{\hat{U}}^{\text{tag'}}_{i,\text{ data}} + \hat{\hat{D}}^{\text{tag'}}_{i,\text{ data}} - (\hat{\hat{U}}^{\text{tag'}}_{i,\ t\bar{t},MC} + \hat{\hat{D}}^{\text{tag'}}_{i,\ t\bar{t},MC}) - (\hat{\hat{U}}^{\text{tag'}}_{i,\text{ V, MC}} + \hat{\hat{D}}^{\text{tag'}}_{i,\text{ V, MC}}) \Big), \tag{7.13}$$

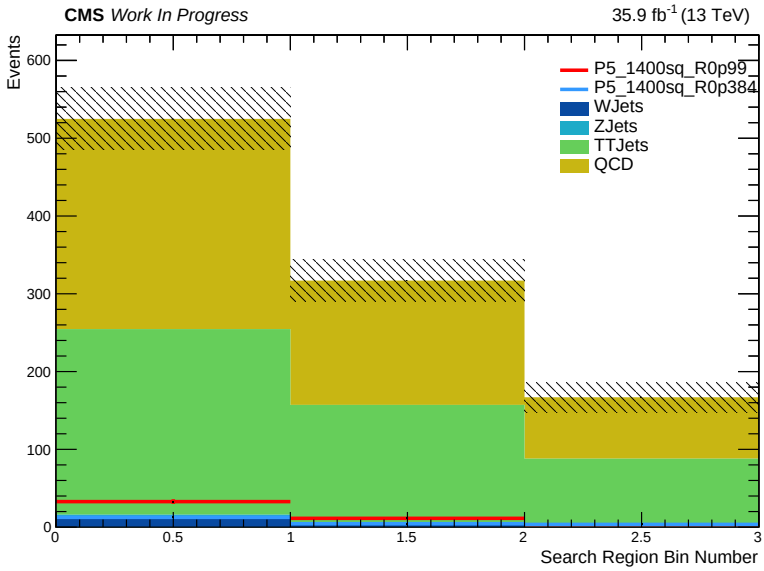
where $\hat{\hat{U}}^{\text{tag'}}_{i,\ t\bar{t},MC}$ and $\hat{\hat{D}}^{\text{tag'}}_{i,\ t\bar{t},MC}$ represent the expected yields from $t\bar{t}$ MC simulated events in the upper and lower sideband bins respectively.

As shown in Equation (7.13), the $t\bar{t}$ data-driven estimated yields in the central bins are influenced by the $t\bar{t}$ MC yields in the sideband bins. However, the $t\bar{t}$ yield is much lower in the sideband bins, and as such any variation in the $t\bar{t}$ MC yield has a very small effect on the data-driven estimation.

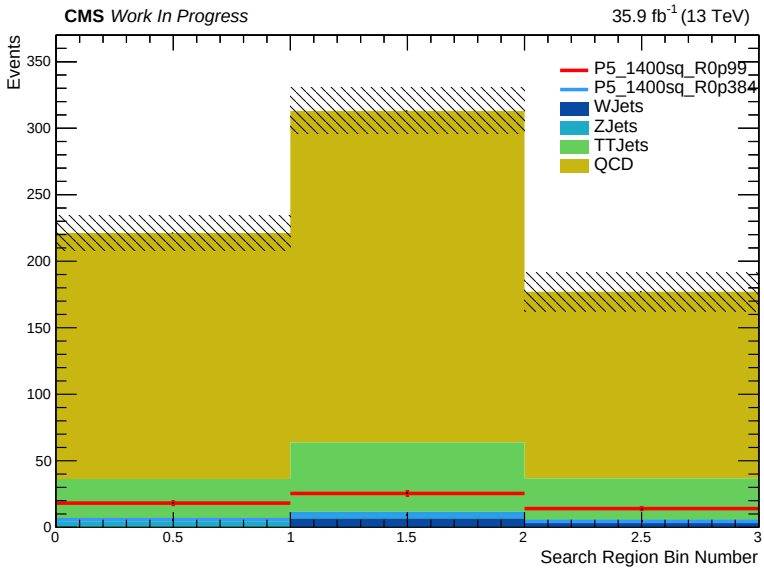
Figure 7.22 shows that in all cases the signal yields are considerably smaller than the background yields, and thus assuming a large systematic uncertainty on the $t\bar{t}$ expected yield this control region appears robust. Additionally we note that QCD is by far the dominant process in the anti-tag control region, allowing for data-driven estimation of its expected yield in the tag' control region. This yield is then input to the calculation of the $t\bar{t}$ expected yield from observation in the U_i and D_i sidebands.

7.6.3 Comparing QCD MC vs Data-Driven Estimation in the $t\bar{t}$-enriched Control Region

In order to calculate data-driven $t\bar{t}$ estimated yields in these control region bins, we first require an estimation for the QCD yields, as shown in Equation (7.13). Firstly, we compare the QCD MC yields to those obtained using the data-driven



(a) Central bands in Tag DBT region.



(b) UnD Sidebands in Tag DBT region.

Figure 7.22: Estimated yields in the central and sideband bins for QCD, $t\bar{t}$, W+Jets and Z+Jets background processes, compared with signal MC benchmark points containing top squark decays.

method in equation 7.11. Here we find the maximum discrepancy between the Monte Carlo simulation and the data-driven estimation method of approximately 15 %, shown in Figure 7.23.

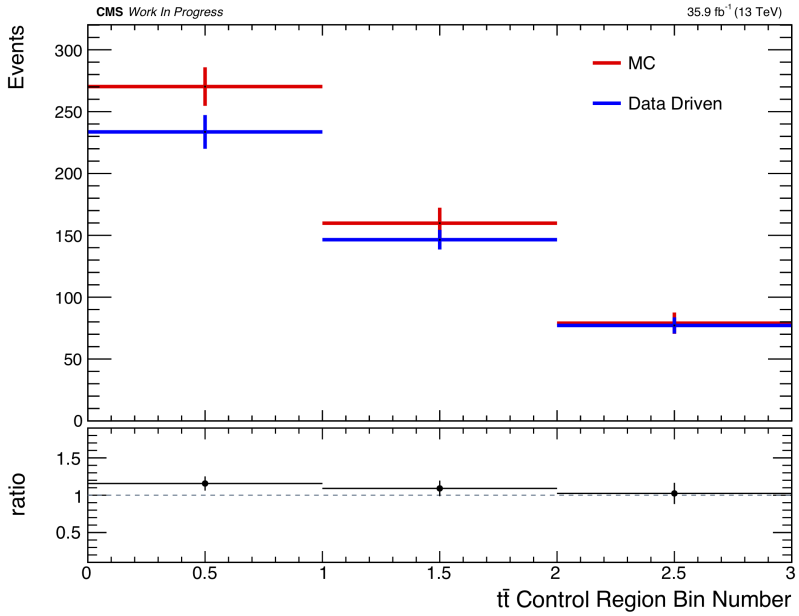
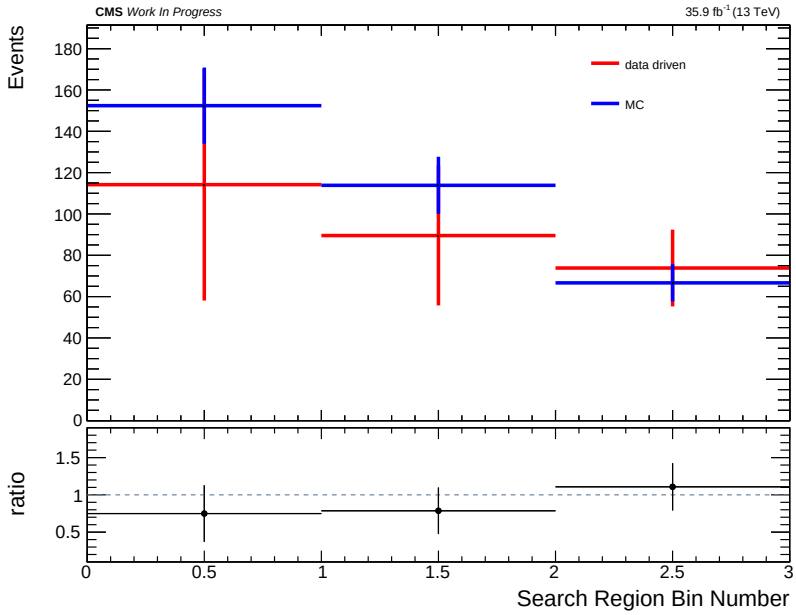


Figure 7.23: QCD MC vs data-driven QCD yields across the control region bins.

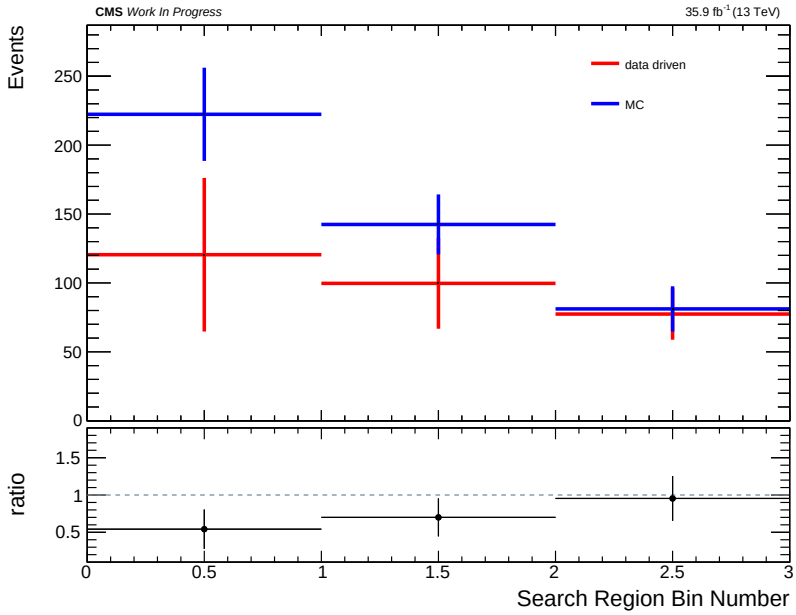
7.6.4 Comparing $t\bar{t}$ MC vs Data-Driven Estimation in the Hadronic CR

Using the value of 15 % motivated in Section 7.6.3 to give an idea of the uncertainty in the estimated QCD yields, we calculate the $t\bar{t}$ expected yields using the data-driven method in Equation (7.10).

We see in Figure 7.24 a comparison between the data-driven $t\bar{t}$ yields and those derived using $t\bar{t}$ MC simulated events using Powheg and MadGraph for run year 2016, with similar results seen for run year 2017. Whilst the discrepancies between data and MC are generally lower for Powheg compared with MadGraph, shown in Figure 7.24, we note that the agreement is generally good considering the respective uncertainties, which correspond to uncertainty in the event yield due to $\pm 1\sigma$ variations in the double-b-tag scale factors, detailed in Section 7.7. Furthermore the MadGraph MC samples are generated such that they contain a much greater



(a) Powheg.



(b) MadGraph.

Figure 7.24: $t\bar{t}$ data-driven estimation compared with yields derived using MC simulated $t\bar{t}$ events.

number of simulated events in high-H_{T} regions, whereas the Powheg samples suffer from zero statistics in a number of signal region bins, thus motivating the use of the MadGraph LO MC $t\bar{t}$ samples, as is demonstrated in Figure 7.25.

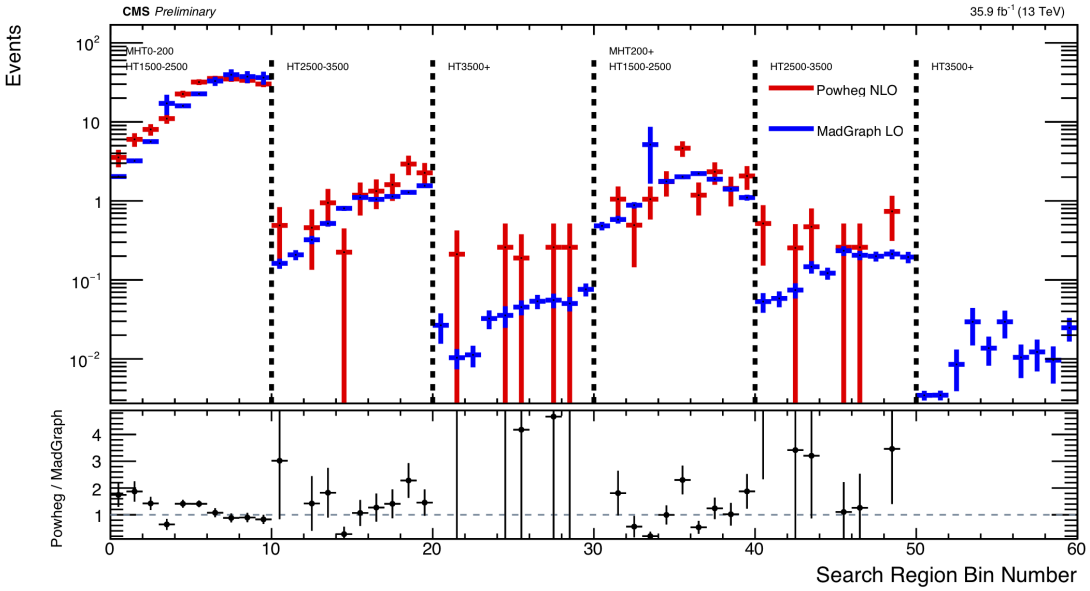


Figure 7.25: Comparison between expected background yields for $t\bar{t}$ events from the Powheg NLO and MadGraph LO samples across the 60 signal region measurement bins in this analysis. We note that many of the bins in the high-H_{T}, high-$H_{\mathrm{T}}^{\mathrm{miss}}$ regions contain zero Powheg events.

We conservatively use the largest discrepancy observed in the first bin of Figure 7.24 as well as the differences between the top quark p_{T} distributions and event yields for the MadGraph LO and Powheg NLO samples, shown in Figures 7.18 and 7.25, to motivate a value of $\sim 100\%$ for the total systematic uncertainty on the $t\bar{t}$ yield. Whilst this value is quite high, we note that for all signal region bins with more than one expected background event, QCD is still the dominant background process, and as such the $t\bar{t}$ uncertainty will not dominate the overall systematic uncertainty on the total background yield.

7.6.5 Conclusions of $t\bar{t}$ Study

On the whole we see a good level of agreement in the top quark p_{T} distributions between the $t\bar{t}$ MC samples generated using Powheg and MadGraph, shown in Figure 7.18, and as such the Powheg-derived top quark p_{T} re-weighting is used for

the MadGraph samples. Furthermore, since only the MadGraph samples contain sufficient statistics in the high-H_{T}, high-$H_{\mathrm{T}}^{\mathrm{miss}}$ regions in this analysis, we use these samples in lieu of those generated using Powheg, albeit with the 100 % systematic uncertainty motivated in Section 7.6.4.

7.7 Systematic Uncertainties

All signal yields in this analysis are calculated from MC simulation, as are the yields for $t\bar{t}$, single top quark and electroweak background processes, with all MC simulated yields containing various sources of systematic uncertainty. The dominant background contribution, from QCD multijet events, is calculated using a data-driven approach, and the systematic uncertainty on the expected QCD event yield is discussed in Section 7.5.

This section details the various sources of systematic uncertainties for each MC simulated process. Variations in H_{T}, $H_{\mathrm{T}}^{\mathrm{miss}}$, AK8 soft-drop mass and double-b-tag score will alter the event yields across the various measurement bins. Since this will of course impact the overall sensitivity, it is therefore important to model the effect of the respective uncertainties which affect these MC simulations.

Since MC simulation is run independently for the 2016 and 2017 run years, this section will also detail any variations in uncertainties between different run years, along with any uncertainties which are correlated between years. Where appropriate, the effects of variations in each systematic uncertainty on the overall expected event yields are shown for an example signal mass point from the BM1 mass scan, with $M_{\tilde{q}} = 1800$ GeV and $M_{\tilde{\chi}_1^0} = 1$ GeV. Furthermore cases where uncertainty variations apply only to a subset of the data, signal MC and background MC samples are indicated as such.

7.7.1 AK8 Jet Double-b-Tag Scale Factors (Signal and $t\bar{t}$ MC)

The double-b-tag score range, extending from -1 to 1, is divided into working points. An AK8 jet is assigned the highest working point which its double-b-tag score surpasses, with the thresholds for loose, medium 1, medium 2 and tight working points being 0.3, 0.6, 0.8 and 0.9 respectively, with all but the tight working point used in this analysis.

Table 7.3: AK8 jet p_{T}-dependent double-b-tag scale factors for signal events for both 2016 and 2017 run years.

AK8 jet p_{T} range	Double-b-tag working point: 0.3 (Loose)	0.6 (Med-1)	0.8 (Med-2)
2016			
250-350 GeV	$0.96^{+0.03}_{-0.02}$	$0.93^{+0.03}_{-0.02}$	$0.92^{+0.03}_{-0.03}$
350-430 GeV	$1.00^{+0.04}_{-0.03}$	$1.01^{+0.03}_{-0.03}$	$1.01^{+0.03}_{-0.04}$
430-840 GeV	$1.01^{+0.02}_{-0.04}$	$0.99^{+0.02}_{-0.04}$	$0.92^{+0.03}_{-0.05}$
840+ GeV	$1.01^{+0.04}_{-0.08}$	$0.99^{+0.04}_{-0.08}$	$0.92^{+0.06}_{-0.10}$
2017			
250-350 GeV	$0.96^{+0.03}_{-0.03}$	$0.93^{+0.04}_{-0.03}$	$0.85^{+0.04}_{-0.04}$
350-840 GeV	$0.95^{+0.06}_{-0.04}$	$0.90^{+0.08}_{-0.04}$	$0.80^{+0.07}_{-0.04}$
840+ GeV	$0.95^{+0.12}_{-0.08}$	$0.90^{+0.16}_{-0.08}$	$0.80^{+0.14}_{-0.08}$

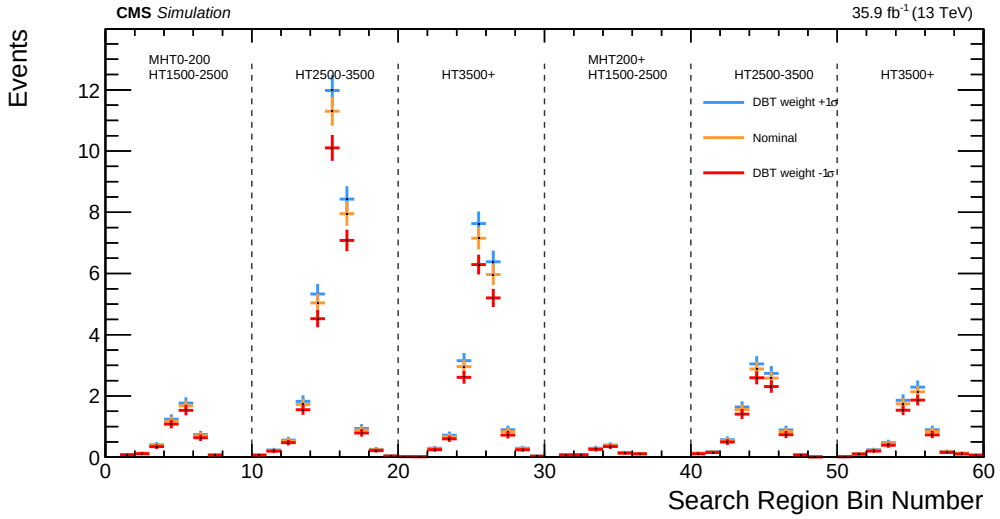
AK8 jet working point- and p_{T}-dependent data and MC scale factors are provided by CMS for run years 2016 and 2017. Furthermore mis-tag scale factors are supplied by CMS for $t\bar{t}$ events where either a top quark or W boson is erroneously reconstructed as an AK8 jet containing two displaced vertices. These respective scale factors are given in Tables 7.3 and 7.4, along with associated systematic uncertainties.

Table 7.4: AK8 jet p_{T}-dependent double-b-tag scale factors for $t\bar{t}$ events for both 2016 and 2017 run years.

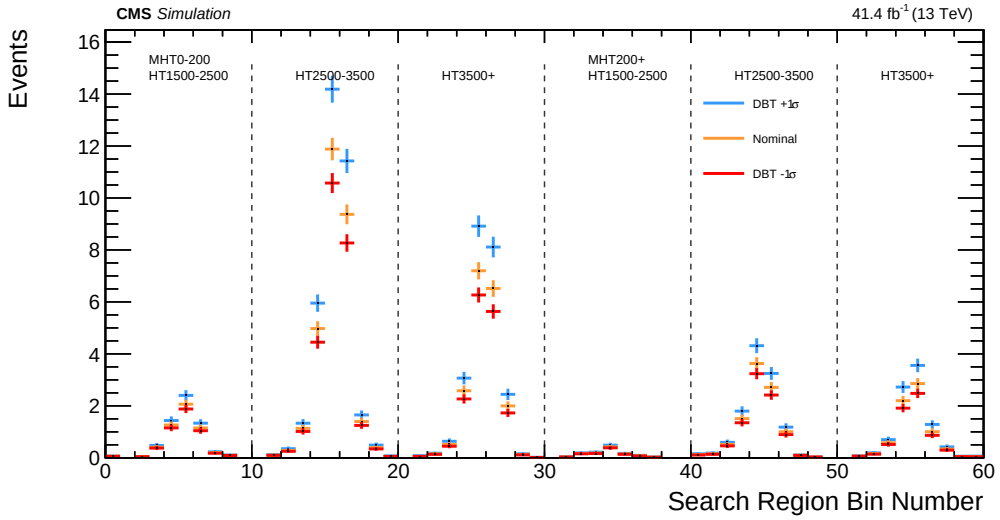
AK8 jet p_{T} range	Double-b-tag working point: 0.3 (Loose)	0.6 (Med-1)	0.8 (Med-2)
2016			
250-350 GeV	1.044±0.028	1.029±0.034	1.050±0.044
350-430 GeV	1.074±0.052	1.156±0.064	1.086±0.078
430-700 GeV	1.119±0.079	1.156±0.064	1.086±0.078
700+ GeV	1.119±0.158	1.156±0.128	1.086±0.156
2017			
250-350 GeV	$0.939^{+0.026}_{-0.026}$	$0.922^{+0.027}_{-0.027}$	$0.875^{+0.030}_{-0.030}$
350-430 GeV	$1.007^{+0.055}_{-0.054}$	$0.967^{+0.057}_{-0.056}$	$0.939^{+0.063}_{-0.063}$
430+ GeV	$0.996^{+0.080}_{-0.078}$	$0.902^{+0.083}_{-0.081}$	$0.893^{+0.091}_{-0.089}$

However, as is shown in Figure 7.6 the signal acceptance region is triangular and thus does not line up exactly with any 2D rectangles constructed from the loose, medium 1 or medium 2 working points. As such an approximation was made in order to evaluate the systematic uncertainty for events falling into the signal region of the 2D double-b-tag score plane.

As shown in Figure 7.26, signal MC in run year 2016 $+1\sigma$ variations in this systematic uncertainty cause a uniform increase of $\sim 10\%$ in yield across all signal bins, with -1σ variations leading to a decrease of $\sim 15\%$. Owing to updates performed to the CMS pixel detector during the 2016 year-end technical shutdown (YETS) and since the double-b-tag BDT was re-trained prior to 2017 data taking. As a result, this effect of $+1\sigma$ variation is larger for 2017, leading to an increase of $\sim 25\%$, and these uncertainties are treated as uncorrelated between run years 2016 and 2017.



(a) Run year 2016.



(b) Run year 2017.

Figure 7.26: Effects of double-b-tag scale factor $\pm 1\sigma$ variations on the overall event yields for an example signal mass point.

7.7.2 AK8 Jet soft-drop Mass

The soft-drop masses of the two AK8 jets leading in double-b-tag score each have two associated uncertainties: An uncertainty for the soft-drop mass scale (JMS) and an uncertainty for the soft-drop mass resolution (JMR), applied to all MC simulated events.

Following the recommendation of CMS, the JMS uncertainty is applied in the form of scale factors representing $\pm 1\sigma$ shifts in the soft-drop mass of each of the two AK8 jets leading in double-b-tag score. The scale factors are provided directly by CMS for 2016 MC, with the recommendation being to use the same factors for 2017 MC; with these uncertainties therefore treated as being correlated between years.

Given there are ten signal region bins in the 2D AK8 soft-drop mass plane for each H_{T}-$H_{\mathrm{T}}^{\mathrm{miss}}$region in this analysis, these variations will naturally lead to variations in the distributions for signal and background processes across the sixty bins. This uncertainty is therefore taken into account for all MC simulated events as a nuisance parameter in the log likelihood function when performing the overall fit and calculating the limits.

However as can be seen for an example signal mass point in Figure 7.27, the overall impact of $\pm 1\sigma$ variations in the JMS scale factors have lesser overall effect on signal MC distributions compared with those for the double-b-tag scale factors.

The remaining soft-drop mass uncertainty from JMR is implemented via smearing applied to the soft-drop mass value of each jet, worsening the resolution. Once again following the CMS recommendation, this smearing value, ΔM, added to the soft-drop mass, such that $M \to M + \Delta M$, is given in Equation (7.14):

$$\Delta M = \mathcal{N}(\sigma_{\mathrm{JMR}}) \cdot \sqrt{s_{\mathrm{JMR}}^2 - 1}, \tag{7.14}$$

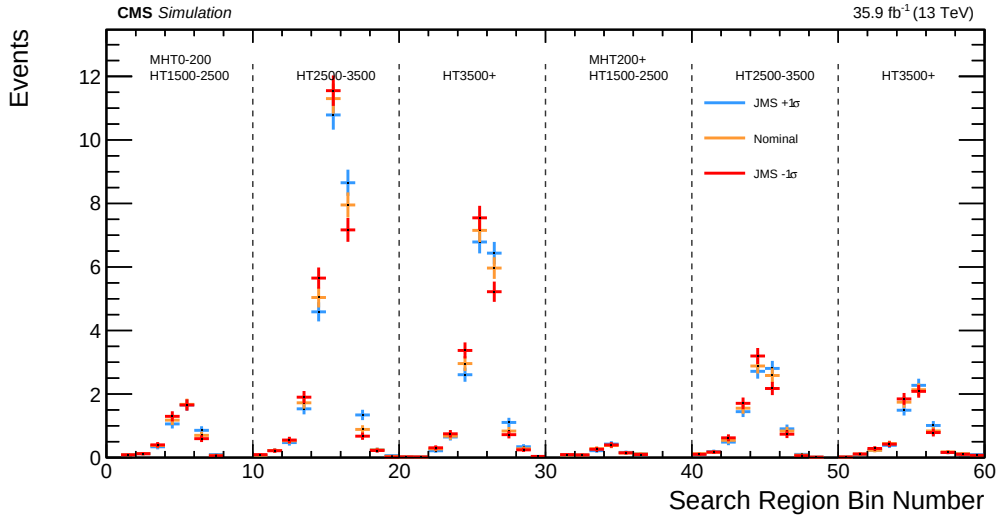


Figure 7.27: Expected yields with $\pm 1\sigma$ variations in JMS scale factors for an example signal mass point.

where $\mathcal{N}$ is a random number from a Gaussian with mean zero and standard deviation $\sigma_{\text{JMR}} = 10.1$ GeV and $s_{\text{JMR}} = 1.20$ is the soft-drop mass resolution scale factor recommended by CMS.

This soft-drop mass resolution uncertainty has the largest effect of all systematic uncertainties in [2], especially for light scalar Higgs bosons with mass $M_H < 70$ GeV. However for an SM-like Higgs boson of mass 125 GeV, as shown in Figure 7.9, the soft-drop mass distribution in the 2D plane is quite broad, compared with that for a light scalar Higgs boson of mass $M_H < 70$ GeV. As a result, for heavier Higgs boson masses the size of the smearing becomes less significant compared to the width of the primary mass distribution. Therefore in this SM-like Higgs boson-focused extension the effect of the JMR uncertainty is relatively much smaller, typically less than 10%, as can be seen in Figure 7.28.

7.7.3 Jet Energy Corrections

Jet energy corrections (JECs) are applied to the p_{T} of both AK4 and AK8 jets in this analysis. This has an associated uncertainty due to uncertainty propagated through from the corrections to each jet, which is applied to all MC simulated events.

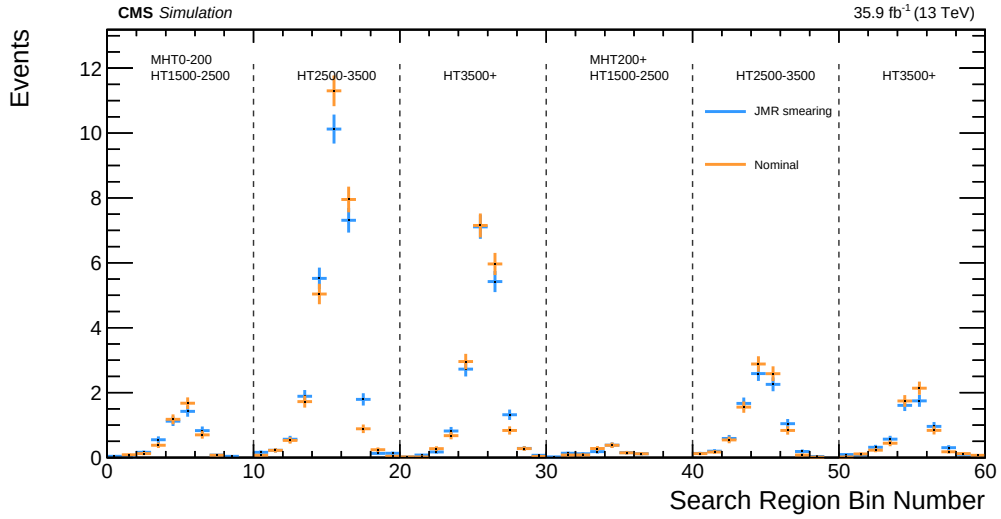


Figure 7.28: Expected yields with and without Gaussian smearing applied to the soft-drop mass resolution, for an example signal mass point.

These uncertainties, recommended by CMS, are correlated between AK4 and AK8 jets since the same corrections are applied to both. Furthermore the uncertainties are also taken to be fully correlated between run years 2016 and 2017, meaning the associated nuisance parameters feature as a joint distribution in the log likelihood function. This has the effect of increasing the uncertainty, since the nuisance parameters corresponding to each run year will be constrained.

Variations in these JECs will impact quantities such as H_{T}, $H_{\mathrm{T}}^{\mathrm{miss}}$, lead AK4 jet p_{T} and AK8 jet p_{T}, thus affecting event yields. These variations are taken into account, with uncertainties propagating to all affected variables. However since the requirement for AK4 and AK8 jet p_{T} is 300 GeV, and most signal events will have jet p_{T} far greater than this threshold, the overall impact on signal yields is small, as is shown in Figure 7.29.

As well as JECs, the p_{T} of AK4 and AK8 jets is also affected by jet energy resolution (JER). The resolution in MC simulation is better than that in data, meaning that jet energy in MC events must have smearing applied in order to better represent the data. Following the CMS recommendation once more, this smearing is applied to each jet in all MC simulated events, with associated uncertainty correlated between AK4 and AK8 jets but in this case not correlated between run years

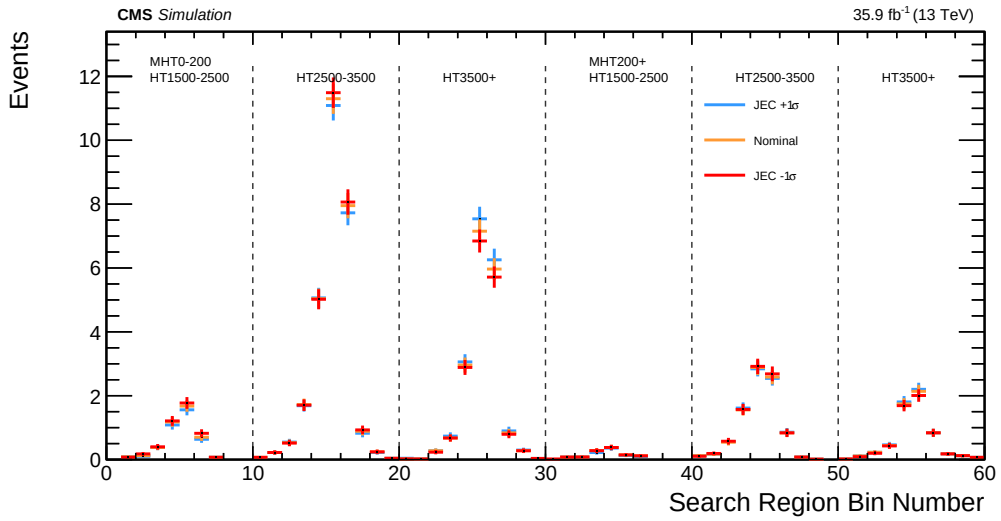


Figure 7.29: Expected yields with $\pm 1\sigma$ variations in JECs for an example signal mass point.

2016 and 2017. As shown in Figure 7.30, the effect of these JER uncertainties on an example signal mass point is vanishingly small

7.7.4 Level-1 Trigger Prefiring

As was detailed in Chapter 3 the Level-1 trigger suffered from a prefiring issue across both the 2016 and 2017 run years. This meant a possibility for events featuring high ECAL deposits could erroneously trigger the previous bunch crossing, meaning the event itself would be missed.

In order to address this issue an event re-weighting is used across all MC simulated events in this analysis. This re-weighting factor, ω, represents the probability that a given event would *not* have caused a trigger prefire had it occurred in the data.

The definition of this re-weighting factor is as follows: All offline photons and AK4 jets in an event are analysed in turn, with a probability assigned to each that they would not cause the prefiring of the Level-1 trigger. The overall factor is then the product of all of these probabilities, shown in Equation (7.15) below:

$$\omega := \prod_{\text{photons, jets}} \left(1 - p^{\text{prefire}}(\eta, p_{\text{T}}^{\text{ECAL}})\right) \tag{7.15}$$

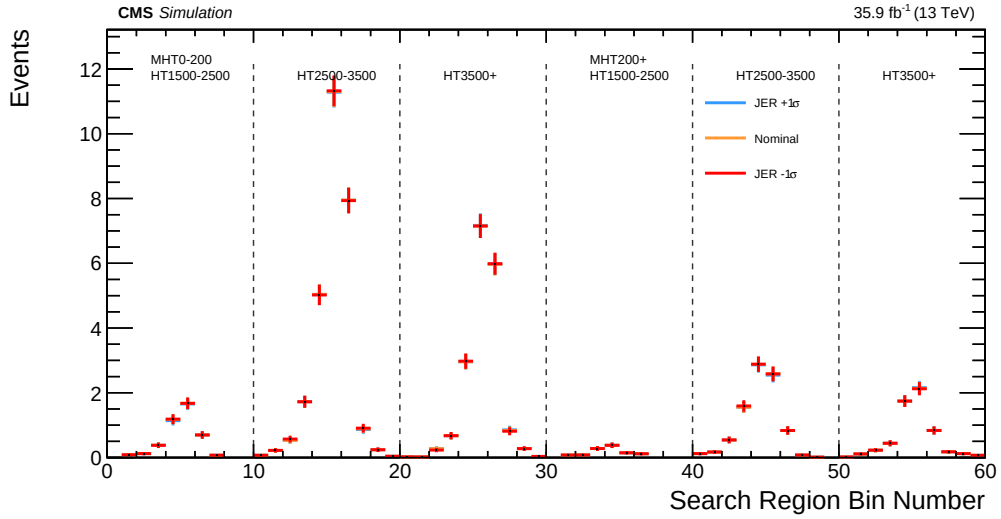


Figure 7.30: Expected yields with $\pm 1\sigma$ variations from JER smearing for an example signal mass point.

This re-weighting factor was calculated using a small subset of 2016 and 2017 data containing only events which could not have been affected by prefiring. This was guaranteed by taking only events which were triggered exactly three bunch crossings after another triggered event. Since the triggering of an event means the next two events cannot be triggered upon, the third event *cannot* self-veto, and is thus immune from prefiring.

The prefiring issue affected mostly events with forward jets, with high $|\eta|$. Whilst the signal MC events feature high-p_{T} jets, these jets are not usually forward-oriented since p_{T} is measured in the plane perpendicular to the beam axis. As such, the overall effect of re-weighting for 2017 signal MC events is only a few percent. Furthermore the re-weighting factors for 2016 signal MC events lie closer to unity, since the prefiring issue worsened from the beginning of 2016 to the end of the 2017 run year due to further deteriorations in the ECAL.

The effect for $t\bar{t}$ MC events was found to be slightly larger than that for signal MC, but follows the same overall shape. The overall average prefire re-weighting factors for an example signal mass point as well as $t\bar{t}$ for both 2016 and 2017 run years are shown in Table 7.5 below.

A systematic uncertainty is attached to this re-weighting factor, and is defined as the maximum of the statistical uncertainty and 20 % of the prefiring probability. This uncertainty is taken to be uncorrelated between 2016 and 2017 run years. The uncertainty has a < 1 % impact on the average prefire re-weighting factors and thus the overall effect on the signal yields, shown in Figure 7.31, is very small.

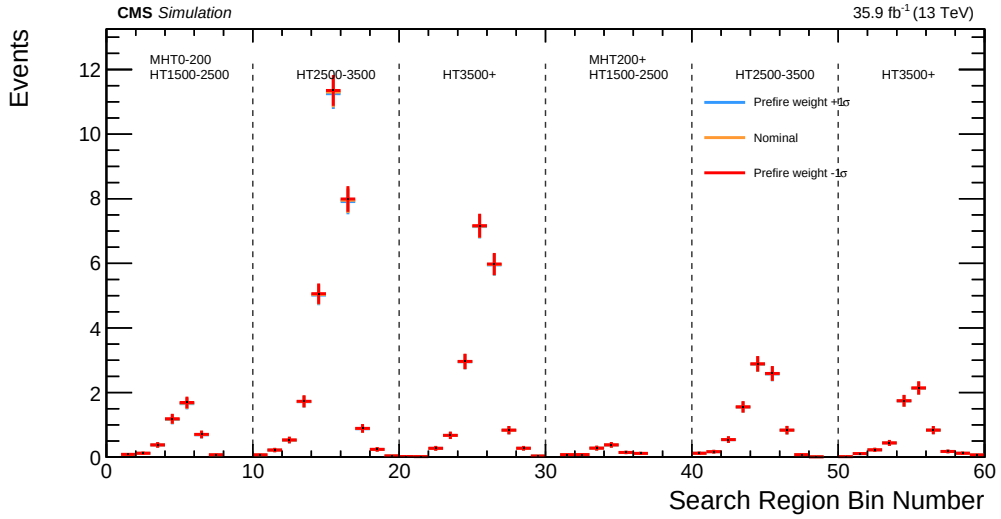


Figure 7.31: Expected yields with $\pm 1\sigma$ variations in prefire scale factor weighting for an example signal mass point.

7.7.5 Luminosity

The luminosity of data recorded by CMS for run years 2016 and 2017 features an uncertainty which is applied to all MC samples, for both signal and background processes. This uncertainty is 2.5% for 2016 and 2.3% for 2017, and was found to have negligible effect on the overall expected limits.

7.7.6 Monte Carlo Event Statistics

Given the very high minimum H_{T} requirement as well as AK8 jet double-b-tagging and soft-drop mass requirements, it is not surprising that in the signal region bins

	2016	2017
BM1: $M_{\tilde{q}} = 1800$ GeV, $M_{\tilde{\chi}_1^0} = 1$ GeV	0.9754	0.96
$t\bar{t}$	0.9665	0.9484

Table 7.5: Average prefire weight for signal and $t\bar{t}$ MC samples in 2016 and 2017.

for which $H_{\mathrm{T}} > 3500$ GeV and $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV, many background events do not pass the event selection criteria.

Signal MC samples generally do not suffer as much from low statistics due to better acceptance, however some background processes feature larger statistical uncertainty. However for the regions where this uncertainty becomes large the overall expected yield is very small, on a sub-event level. As is detailed in Section 7.6, $t\bar{t}$ samples generated with high statistics in high-H_{T} regions were used in order to give non-zero event yields across the analysis binning.

As such, whilst the systematic uncertainty is considered for both signal and background MC samples when calculating expected and observed limits, its effect is very small.

7.7.7 Background Monte Carlo Event Yield Uncertainty

The existing analysis in Ref [2] applied a flat 50 % uncertainty on the total event yields for $t\bar{t}$, Z+jets and W+jets for both the 2016 and 2017 run years in order to account for uncertainties in MC event generation, noting that these background processes are sub-dominant.

In this analysis the $t\bar{t}$ background modelling has been extensively studied in Section 7.6, which motivated a systematic uncertainty of 100 % for the total number of expected $t\bar{t}$ events. For the remaining single top quark and electroweak background processes the estimated value of 50 % is kept, once again noting that these background processes are entirely sub-dominant, with the overall systematic uncertainty on the respective event yields having negligible effect on the overall expected background yields in this analysis.

7.7.8 Veto on Isolated Leptons, Photons and Tracks

The object veto in this analysis applied to events containing isolated leptons, photons and tracks. However owing to the kinematic event selection strongly favouring

fully-hadronic events, this veto applies to very few background MC events and essentially no signal MC events.

Therefore it is taken that any uncertainty originating from this lepton veto will have little to no effect on the overall expected limit, and in the case of background MC event yields is more than accounted for by the conservative choices of cross-section uncertainty for these background processes.

7.7.9 Initial State Radiation Reweighting (Signal MC)

Initial State Radiation (ISR) occurs when radiated jets stemming from the initial state partons enter an event reconstruction. Since these do not originate from any vertex following the pp interaction, but may still enter the calculation of variables such as H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$, this phenomenon must be taken into account for signal MC events.

The correction for ISR is derived in Ref [2] from $t\bar{t}$ events decaying fully leptonically. Requiring two leptons (electrons or muons) and two b-tagged AK4 jets implies that any remaining jets will likely have radiated from the initial state.

These correction factors were found to be 1.000, 0.920, 0.821, 0.715, 0.662, 0.561 and 0.511 for events with numbers of ISR jets $n_{\mathrm{ISR}} = 0, 1, 2, 3, 4, 5$ and 6+ respectively. These factors were then applied as re-weighting factors to the signal MC events, normalised such that the overall cross-section is unaffected, with the $\pm 1\sigma$ systematic uncertainty value on each factor taken to be half of the difference between the respective value and unity. The overall effects of the ISR re-weighting factors on the event yields across the binning in this analysis is very small, around 2%, as illustrated in Figure 7.32 for an example signal mass point.

7.7.10 Scale Weighting (Signal and $t\bar{t}$ MC)

The QCD scale uncertainty accounts for missing higher orders in the QCD calculations when estimating the production cross-section of a process. Varying the

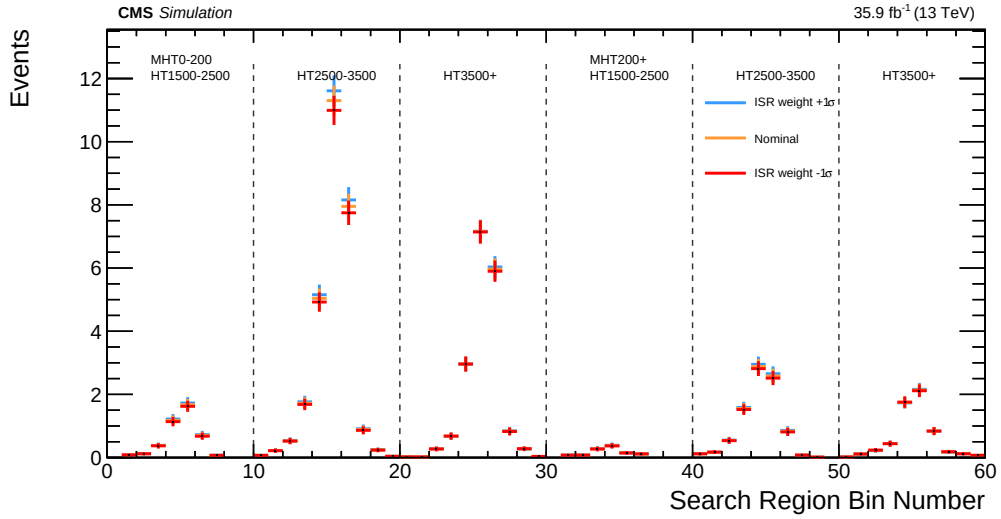


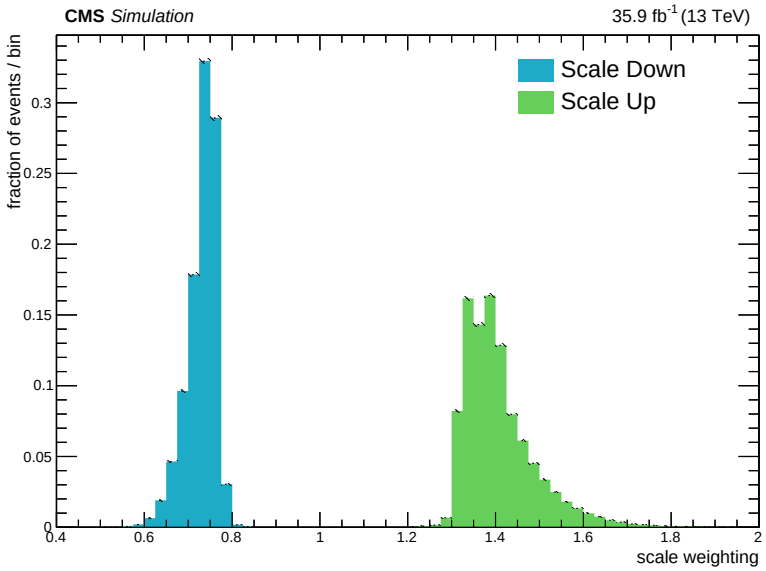
Figure 7.32: Expected yields with $\pm 1\sigma$ variations in ISR re-weighting factors for an example signal mass point.

renormalisation and factorisation scales, μ_{R} and μ_{F}, in the perturbation theory calculations affects the overall cross-section for signal and background process initial state production. The effect of this was investigated for MC simulated NMSSM signal events specific to the analysis in Ref [2].

Since these signal events were simulated using the exact same framework and analysis tools as the signal models considered in this interpretation, and as the initial squark and gluino production will affect this scale uncertainty but the Higgs boson mass will not, since direct Higgs boson production is not calculated at matrix element level in this analysis, it is reasonable to omit the repetition of this investigation with different signal mass points for this interpretation.

The $+1\sigma$ and -1σ scale weights for events correspond to varying these scale factors up and down by a factor of 2; the cases where $\mu_{\mathrm{R}} = \mu_{\mathrm{F}} = 0.5$ and $\mu_{\mathrm{R}} = \mu_{\mathrm{F}} = 2$ respectively. These $\pm 1\sigma$ event weights for an example signal mass points are shown in Figure 7.33, with light scalar Higgs boson mass 70 GeV and squark masses of 1200, 2000 and 2600 GeV respectively, where the nominal event weight is unity.

We see in Figure 7.33 a shift of approximately $-25/+40\%$ from the nominal value. Furthermore it was found that the scale weighting does not change dramatically as the squark mass increases from 1200 GeVto 2600 GeV, despite this large differ-



(a) $M_{\tilde{q}} = 2000$ GeV.

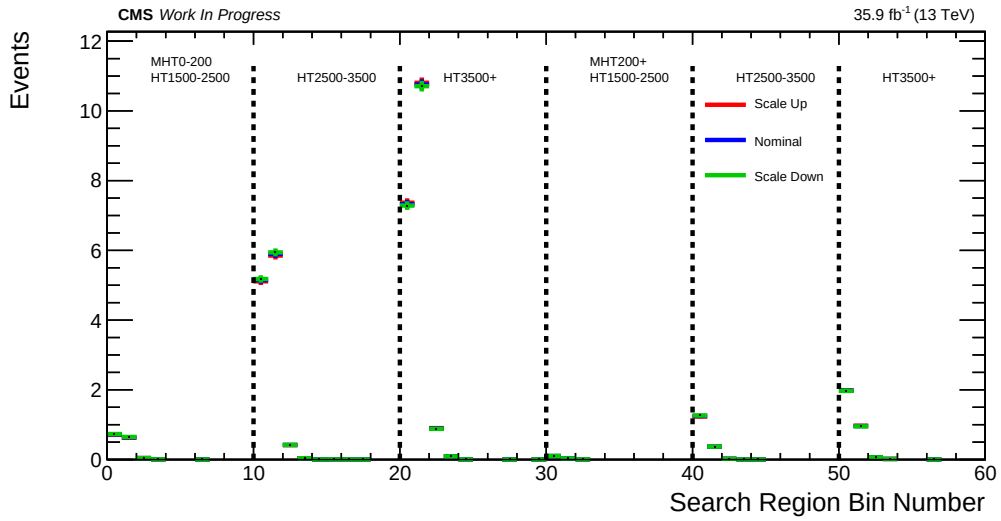
Figure 7.33: Scale weights $\pm 1\sigma$ for an example signal mass points with squark mass 2000 GeV.

ence in mass which corresponds to a difference of multiple orders of magnitude in production cross-section.

These up and down scale weights are normalised by dividing by the average of the $\pm 1\sigma$ scale weighting distributions, before being applied to each event, in order to examine the effect of such variations in μ_{R} and μ_{F} whilst keeping the production cross-section constant. Figure 7.34 shows the change in shape across the analysis binning resulting from normalised $+1\sigma$ and -1σ variations in the scale weighting.

We see that the overall effect on the shape of the event yields in the signal bins is very small, in most cases of the order of a few percent. Therefore the effect on the overall signal sensitivity may be taken into account within the uncertainty on the production cross-section; a flat systematic uncertainty with a conservative value of 25 % on the overall event yields.

For MC simulated $t\bar{t}$ background events, this $\pm 1\sigma$ uncertainty in the scale weighting is included directly as a systematic uncertainty in the overall fit when calculating the limits. Figures 7.35 and 7.36 show respectively the distribution of these



(a) $M_{\tilde{q}} = 2000$ GeV.

Figure 7.34: Shapes of the (normalised) event yield distribution with $\pm 1\sigma$ variations in scale factor weights for an example signal mass point with squark mass 2000 GeV.

scale weights and their effect on the shape of the overall signal region yields for $t\bar{t}$ background.

For the remaining single top quark production and electroweak background processes, the event yields were found to increase and decrease by around 20 % with $\pm 1\sigma$ variations in the scale factor weights, with no change to the overall shape across the measurement bins. As such a simple 20 % systematic uncertainty is applied to the yields for these background processes.

7.7.11 Pileup Reweighting

The overall event yields in Ref [2] have very little PU dependence, owing to the very high object p_{T} and total H_{T} requirements in the kinematic event selection. Furthermore, other than vetoing 2017 MC events with zero PU, neither PU reweighting nor its associated uncertainty are applied to the original analysis.

In order to determine the PU-dependence of the signal mass points in this analysis samples were split into two sub-samples depending on the number of PU interactions, n_{PU}. These samples were found to have marginal difference in yield between the two categories, for which $n_{\mathrm{PU}} \leq 23$ and $n_{\mathrm{PU}} \geq 24$ respectively. A difference

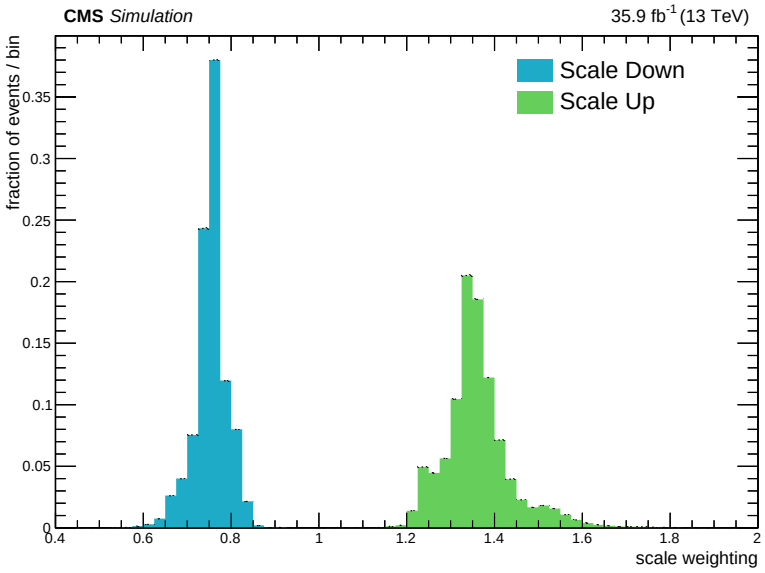


Figure 7.35: Scale weights $\pm 1\sigma$ for $t\bar{t}$.

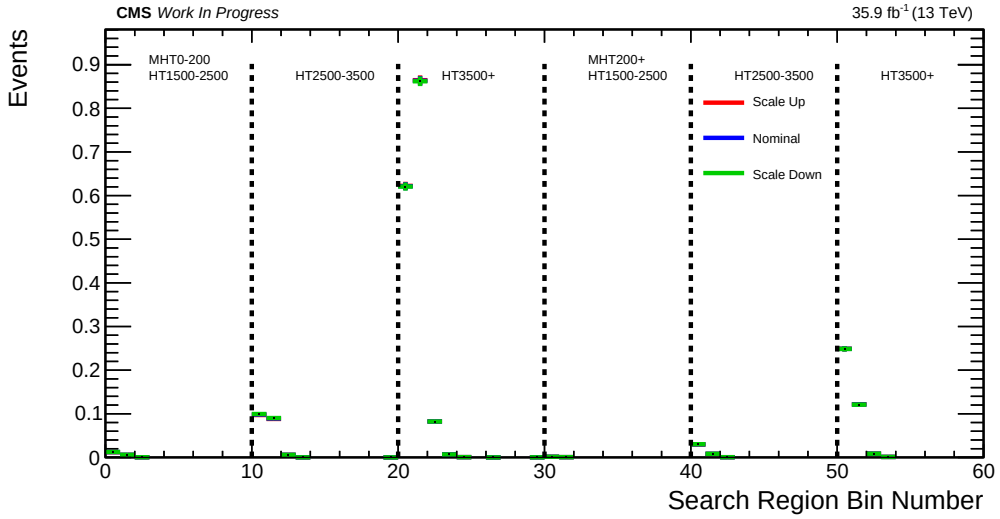


Figure 7.36: Shape of the (normalised) event yield distribution with $\pm 1\sigma$ variations in the scale factor weights for $t\bar{t}$ background.

of this magnitude is well within the uncertainty placed on the signal cross-section, and also dwarfed by other discrepancies which have been shown previously in this section to have an effect of only a few percent on any overall event yields.

Figure 7.37 shows a comparison between events with $n_{\mathrm{PU}} \leq 23$ and $n_{\mathrm{PU}} \geq 24$ for an example signal mass point from the BM1 mass scan, with $M_{\tilde{q}} = 1800$ GeVand $M_{\tilde{\chi}_1^0} = 1$ GeV.

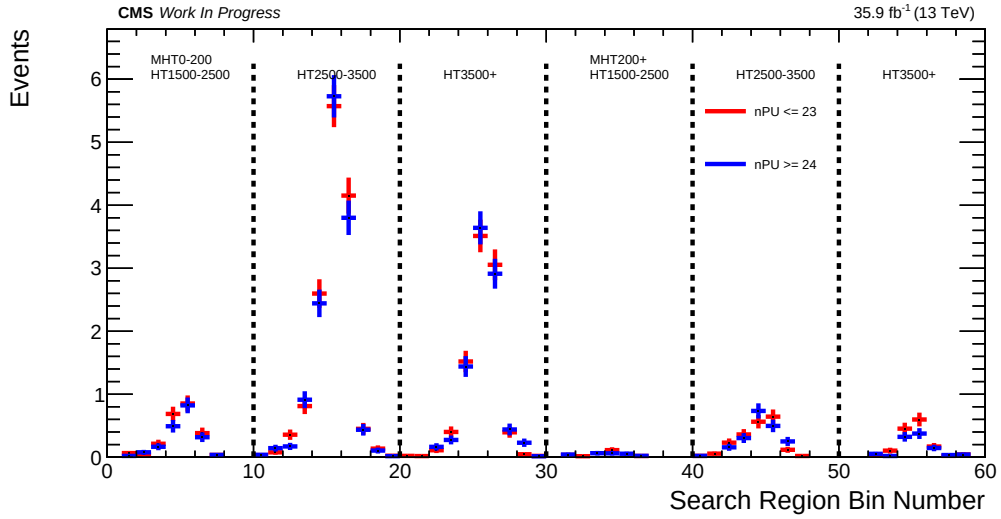


Figure 7.37: Comparison between event yields for $n_{\mathrm{PU}} \leq 23$ and $n_{\mathrm{PU}} \geq 24$ for example mass point with $M_{\tilde{q}} = 1800$ GeV and $M_{\tilde{\chi}_1^0} = 1$ GeV.

7.8 Results

In this section the event yields for signal and background processes are examined, along with observed yields from data, in this analysis, both with and without the inclusion of $H_{\mathrm{T}}^{\mathrm{miss}}$ binning. Expected and observed limits are calculated and explored, with these results compared with those obtained using the unmodified analysis as well as those from the phenomenological work in Chapter 6.

7.8.1 Likelihood Model and QCD Multijet Background Fit

In order to ascertain the level of compatibility between the observed results and the expected SM background yields, a likelihood model is constructed, shown in Equation (7.16). The overall likelihood function comprises the product of separate likelihood functions for the 60 central and 60 sideband measurement bins across each of run years 2016 and 2017. The parameters which characterise this likelihood function are the signal strength, μ, and the nuisance parameters, represented as $\vec{\theta}$.

$$\mathcal{L}(\mu, \vec{\theta}) = \mathcal{L}_{\mathrm{constrain}}(\vec{\theta}) \cdot \prod_{m=S}^{U+D} \cdot \prod_{i=1}^{60} \cdot \prod_{y=2016}^{2017} \mathrm{Poisson}\left(n_{m,i,y} | b_{m,i,y}^{\mathrm{total}}(\vec{\theta}) + \mu \cdot s_{m,i,y} \right), \tag{7.16}$$

where $n_{m,i,y}$ represents the observed number of events in mass region m in measurement bin i for run year y. Similarly, for the same measurement bin $b_{m,i,y}$ and $s_{m,i,y}$ represent the expected total background and signal event yields respectively. This background yield $b_{m,i,y}$ consists of a QCD estimated yield, $b_{m,i,y}^{\mathrm{QCD}}$, and an expected yield for other background contributions calculated using MC simulated events, $b_{m,i,y}^{\mathrm{MC}}$, such that:

$$b_{m,i,y}^{\mathrm{total}} := b_{m,i,y}^{\mathrm{QCD}} + b_{m,i,y}^{\mathrm{MC}}. \tag{7.17}$$

During the fitting procedure the values of the $b_{m,i,y}^{\mathrm{QCD}}$ nuisance parameters, θ, are determined as follows. The QCD yields in the central mass regions S and sideband

mass regions $U + D$ are coupled by factor $F_{i,y}$, which is the F_i factor calculated in Section 7.5 with an extra index y to denote run year, since these F_i factors differ slightly between years:

$$b^{\mathrm{QCD}}_{S,i,y} = F_{i,y} \cdot b^{\mathrm{QCD}}_{U+D,i,y}. \tag{7.18}$$

Each of these $F_{i,y}$ factors is another independent nuisance parameter, with each having its own associated uncertainty values as described in Section 7.5, given a Gaussian probability density function which forms part of the $\mathcal{L}_{\mathrm{constrain}}$ part of the total likelihood model.

These $F_{i,y}$ values are then varied within their associated uncertainty such that the respective $b^{\mathrm{QCD}}_{m,i,y}$ are most compatible with any gap in event yield between the observed yield and the total MC yield from signal and other background processes. In the case of a signal process being present in data in a given measurement bin, the QCD event yield will not lead to double-counting of the expected excess events, since the sidebands were constructed to be essentially signal-free. Figure 7.38 shows a comparison between the relative yields in the central and sideband mass regions for an example signal mass point with $M_{\tilde{q}} = 1800$ GeV and $M_{\tilde{\chi}^0_1} = 1$ GeV.

In the case of certain measurement bins, particularly in regions where $H_{\mathrm{T}} > 3500$ GeV and where both $H_{\mathrm{T}} > 2500$ GeV and $H^{\mathrm{miss}}_{\mathrm{T}} > 200$ GeV, we find that no observed events enter either the respective central or sideband region. This would imply an initial QCD estimation of zero for that particular region, which would cause issues when fitting the particular $F_{i,y}$ value, since this value represents a ratio between the central and sideband regions and so could tend towards infinity should there be zero events in the sideband region. As such in these regions an initial value is predicted using information from the corresponding measurement bin in the anti-tag double-b-tag region.

Generally the number of QCD MC events in a given measurement bin in the tag double-b-tag region is of the order of ~ 0.25 times that in the anti-tag region, with

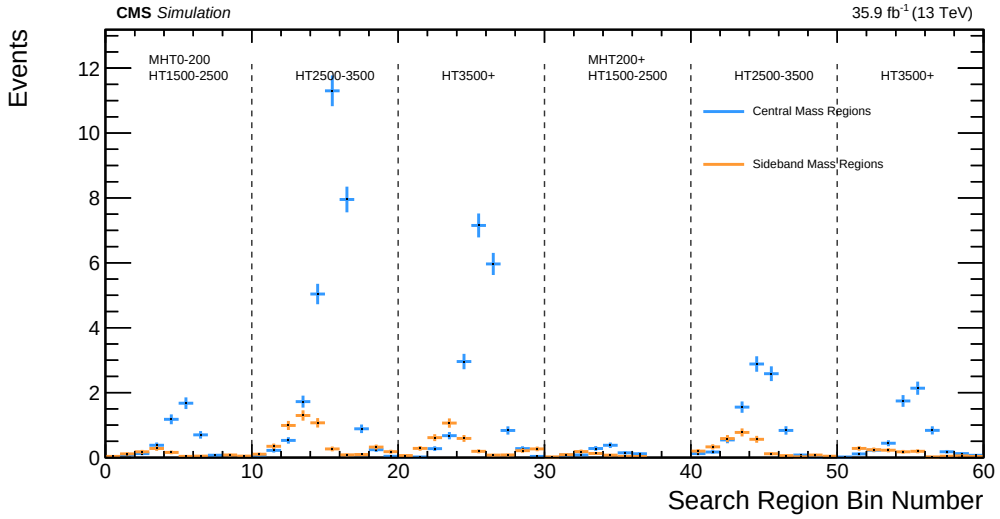


Figure 7.38: Comparison between expected event yields across the 60 central and sideband measurement bins for an example signal mass point in the BM1 mass scan, with $M_{\tilde{q}} = 1800$ GeV and $M_{\tilde{\chi}_1^0} = 1$ GeV.

the MC events having event weight less than unity. As such, in the cases where a sideband measurement bin contains no observed events, the initial QCD estimate in that bin is set to one quarter of the expected yield in the respective anti-tag region bin, implying an initial QCD estimate of this value multiplied by $F_{i,y}$ in the corresponding central bin and noting that the final value of this yield will differ after being allowed to float during the fit.

The MC-derived yields $b^{\text{MC}}_{m,i,y}$ and $s_{m,i,y}$ each have an associated collection of systematic uncertainties, as detailed in Section 7.7. Each of these independent systematic uncertainties is assigned a nuisance parameter, with the nuisance parameters following probability density functions which place constraints on the amount by which the associated parameters may vary from their initial estimated values during the fit and which form the remainder of the $\mathcal{L}_{\text{constrain}}$ term.

It may be noted that in the case of a large discrepancy between the number of observed events in a given central mass bin and its sideband bin, the QCD estimated yield may not be varied infinitely in order to perfectly agree with both observed yields. The reason for this is that starting from the initial estimates for

the QCD yields, derived from observation, it is the F_i ratio values which are then varied, within the constraints set by the associated nuisance parameters.

The systematic uncertainties applied to the MC expected yields are treated as correlated between the mass regions and measurement bins for each given run year, as well as between MC samples where appropriate. Furthermore, as described in Section 7.7, some systematics are considered correlated between run years 2016 and 2017, for example the systematic uncertainty in the JECs. These systematic uncertainties follow a log-normal distribution, due to its robust behaviour in the case where event statistics are low [109].

Moreover, statistical uncertainties are considered for all MC-generated signal and background processes, noting that for the most part the signal statistical uncertainty will be small owing to high event selection efficiencies. In order to model these independent statistical uncertainties the gamma distribution is used, since for MC events the number of generated events is usually greater than the expected event yield, with the expected event yield often taking non-integer values [109].

7.8.2 Expected and Observed Event Yields

Following the full kinematic event selection criteria described in section 7.2, we have sixty event yields corresponding to the measurement bins defined in the signal region. These yields thus comprise number of events which lie in each of the ten central "S" mass regions for each of the two $H_{\mathrm{T}}^{\mathrm{miss}}$ regions within each of the three H_{T} regions.

Figures 7.39 and 7.42 show the distribution of expected background yields for the four combined background processes, along with the predicted yields for three example signal mass points from the BM1 mass scan, coloured blue, orange and red. Similar comparisons for each of the remaining five mass scans may be found in Appendix D.

In Figure 7.40 we see the observed yields across the sixty measurement bins for each run year, compared with pre-fit background yields. Figure 7.41. Similarly Figures 7.43 and 7.44 show the pre- and post-fit yields for run year 2017.

We see in Figures 7.39 and 7.42 for the case of a very light ~ 1 GeV LSP, coloured blue, a signal presence would imply an excess in data above the expected background yield for bins in the highest two H_{T} regions, for both $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV and $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV regions. As $M_{\tilde{\chi}_1^0}$ increases so does the expected $H_{\mathrm{T}}^{\mathrm{miss}}$ distribution, with the average H_{T} decreasing; this pushes such events containing heavier LSPs mostly toward the mid-range $H_{\mathrm{T}} \in [2500, 3500)$ GeV region within the highest $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV region, corresponding to bins 41-50.

As the LSP mass reaches and exceeds 200 GeV the mean H_{T} decreases to the point where most events lie in the lowest $H_{\mathrm{T}} \in [1500, 2500)$ GeV range, where the expected background yields are much higher, thus decreasing sensitivity. Most events from mass points where $M_{\tilde{\chi}_1^0}$ is much higher than this would not be expected to pass the kinematic event selection, since the heavier LSP would inherit most of the NLSP momentum, giving non-boosted Higgs bosons and thus reducing the number of merged AK8 double-b-tagged jets.

The pre-fit background yields in Figures 7.40 and 7.43 generally show good agreement between background estimation and observed event yields. In the case of some of the measurement bins containing few or no observed events we note that the QCD initial estimation is often very small; this occurs when there are no observed events in the sideband regions for a given bin, for example bin 60 in Figure 7.46a. In these cases the initial QCD sideband yield is assigned a small, non-zero initial value, as detailed in Section 7.8.1.

Having ascertained initial non-zero estimates for the QCD sideband yields, such that the F_i factors may be varied during the fit in order to best approximate the sideband and central bin QCD yields, the post-fit yields for the background-only hypothesis are shown in Figures 7.41 and 7.44 for run years 2016 and 2017

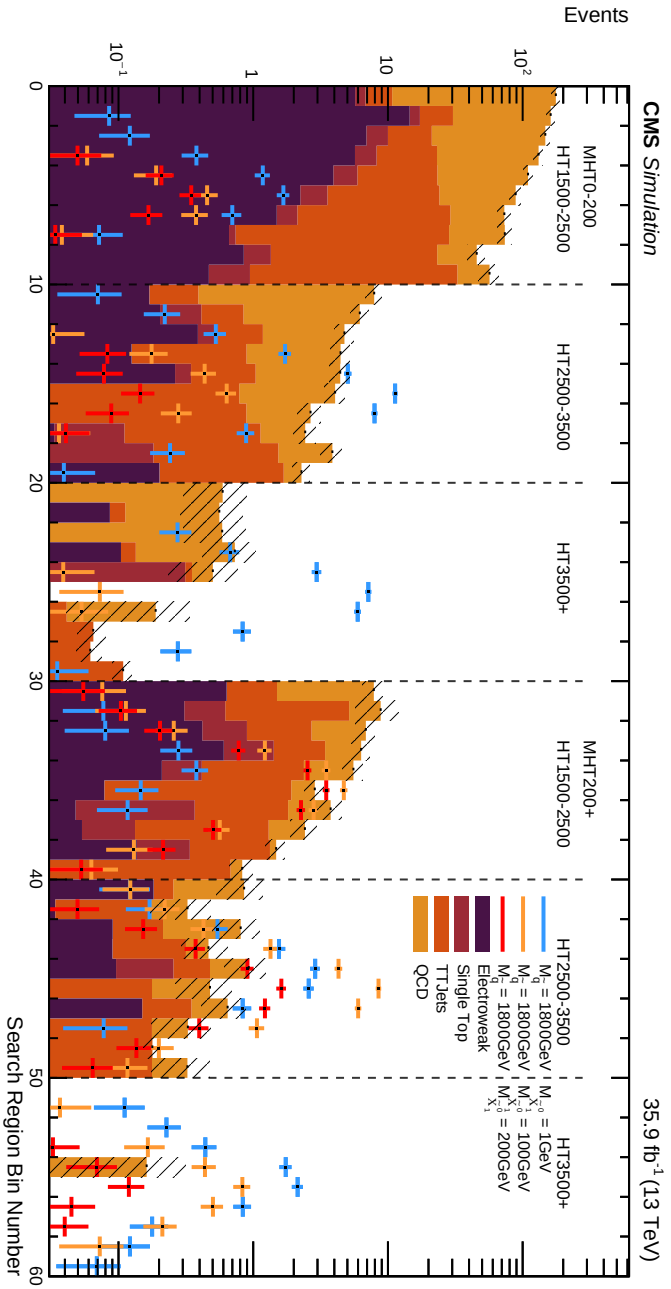


Figure 7.39: MC expected yields across the 60 measurement bins for three example signal mass points from the BM1 mass scan compared with SM MC background processes for run year 2016.

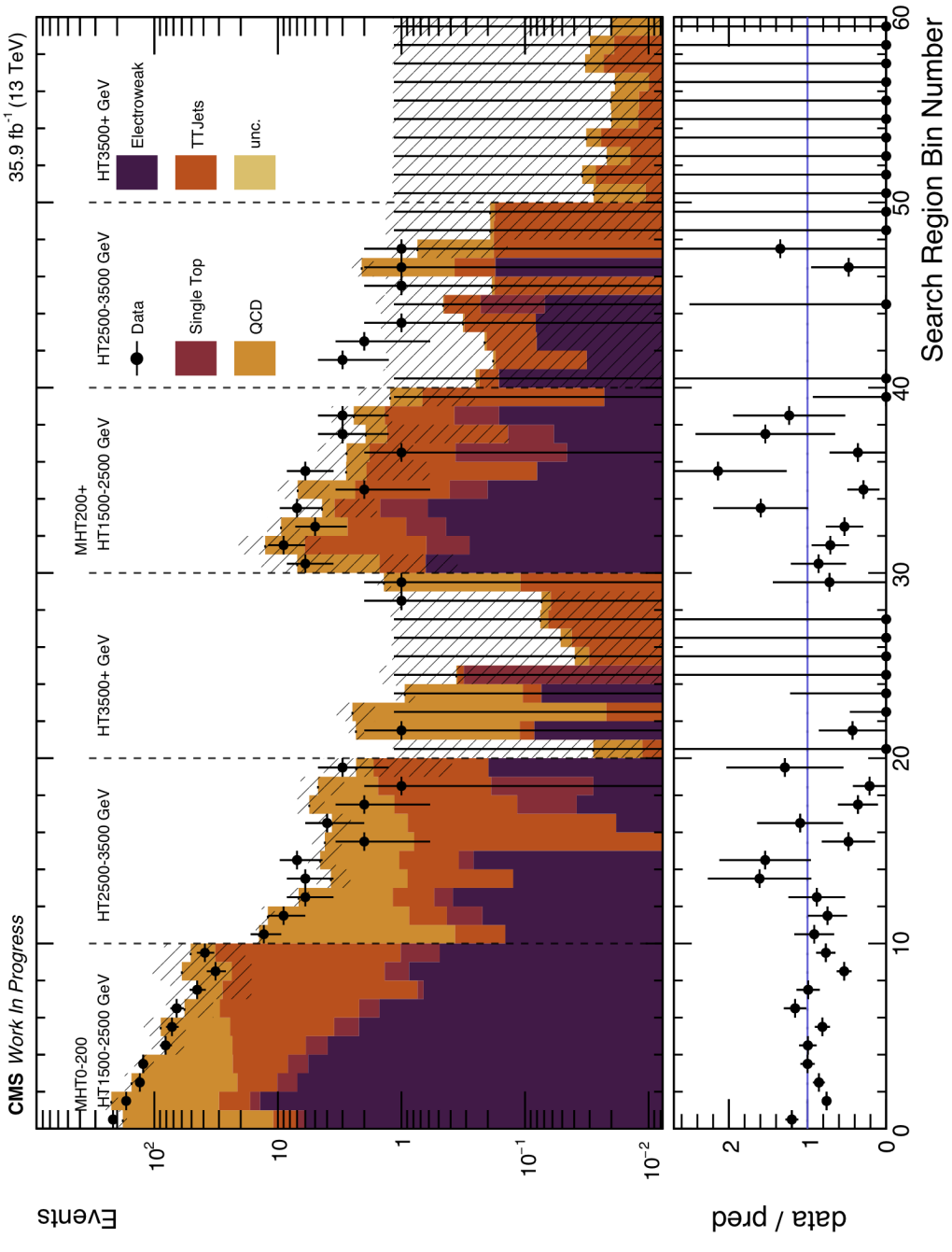


Figure 7.40: Observed and pre-fit background yields across the 60 measurement bins for run year 2016.

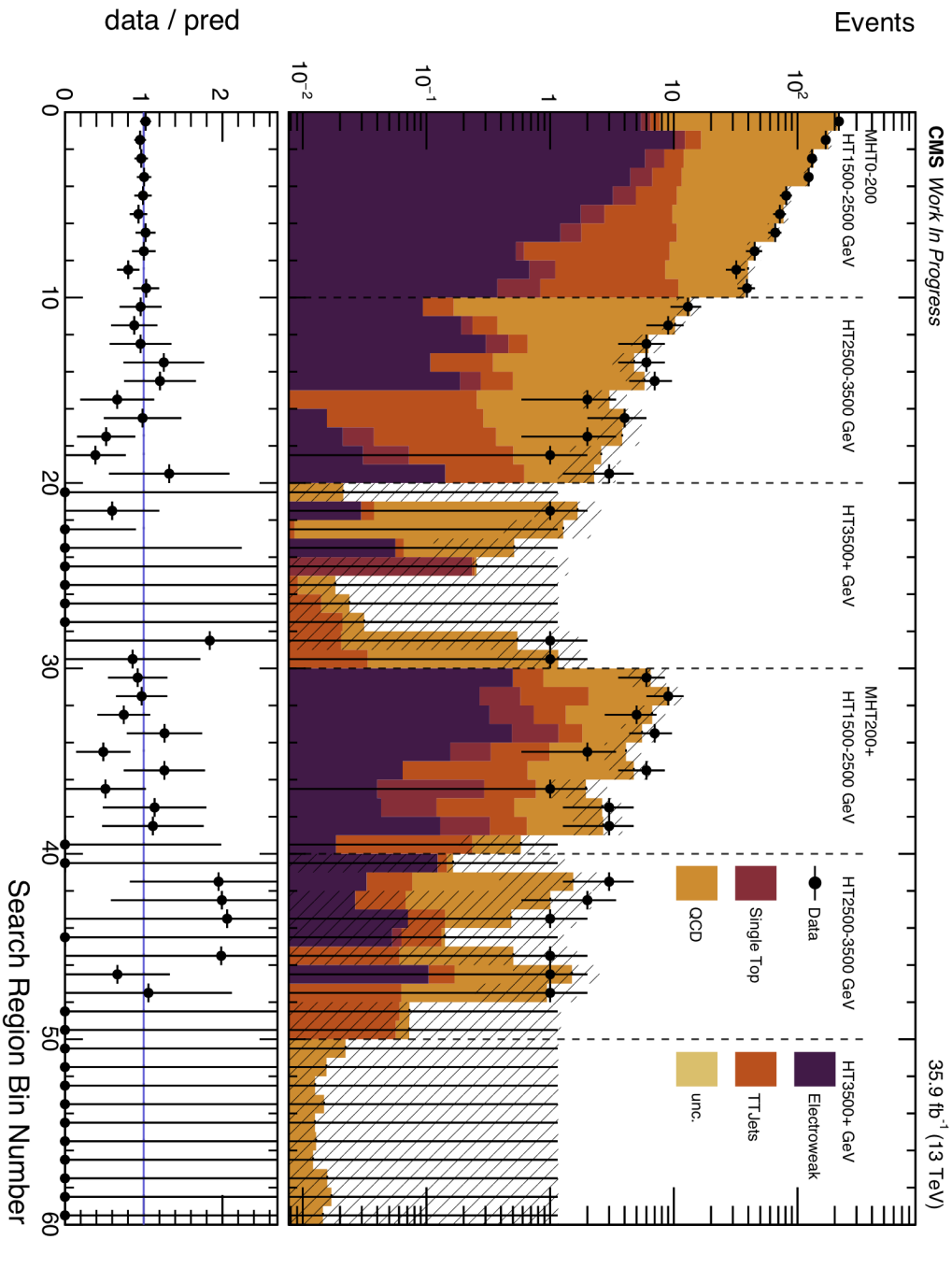


Figure 7.41: Observed and post-fit background yields across the 60 measurement bins for run year 2016.

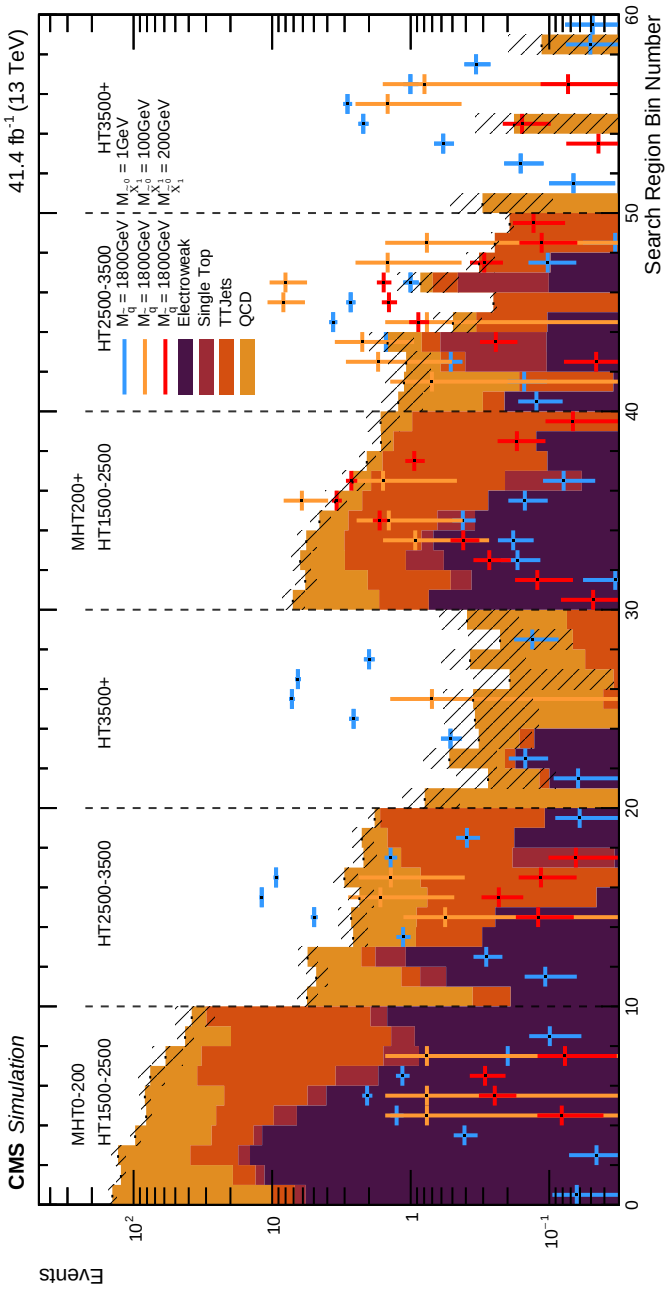


Figure 7.42: MC expected yields across the 60 measurement bins for three example signal mass points from the BM1 mass scan compared with SM MC background processes for run year 2017.

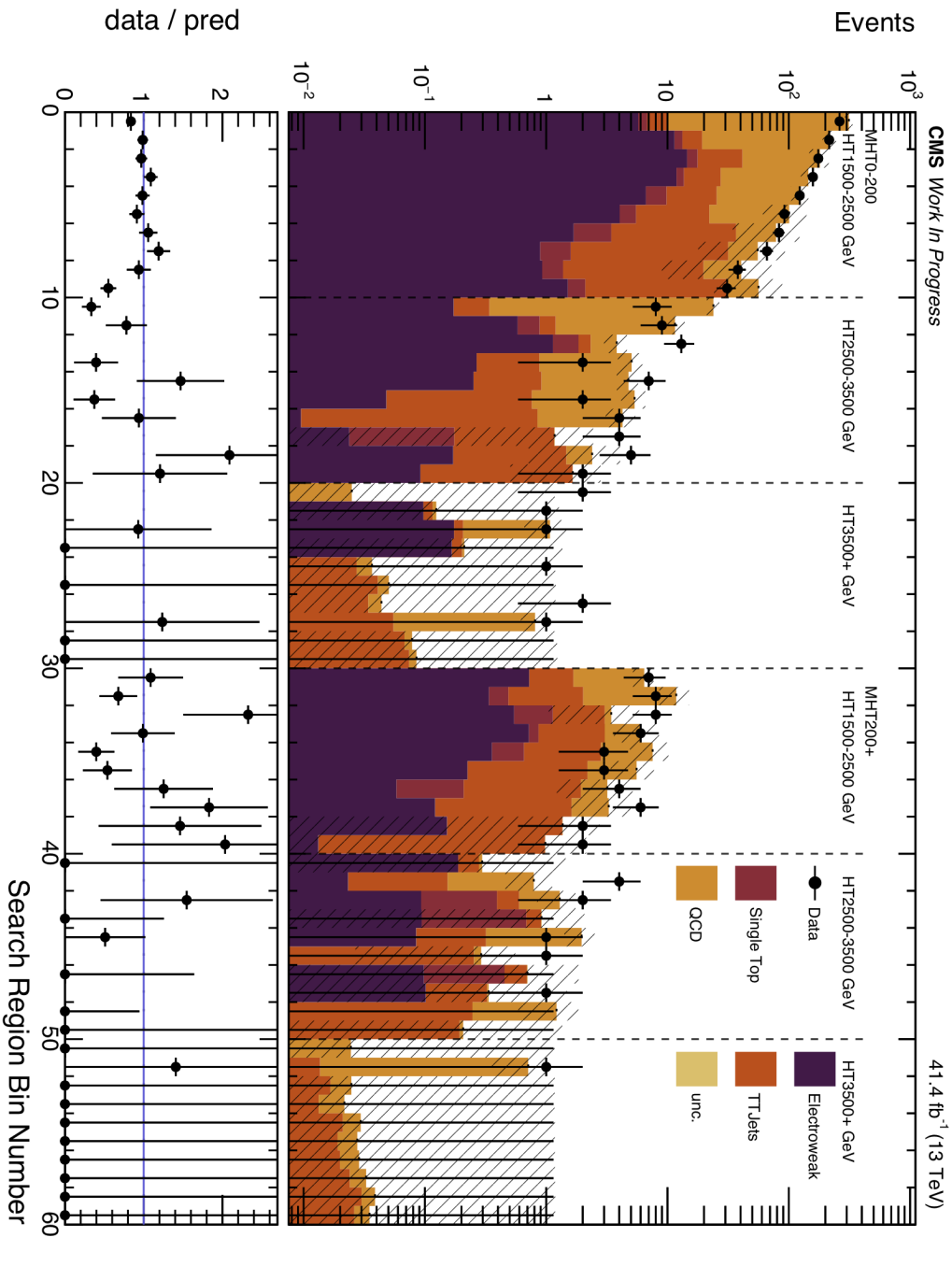


Figure 7.43: Observed and pre-fit background yields across the 60 measurement bins for run year 2017.

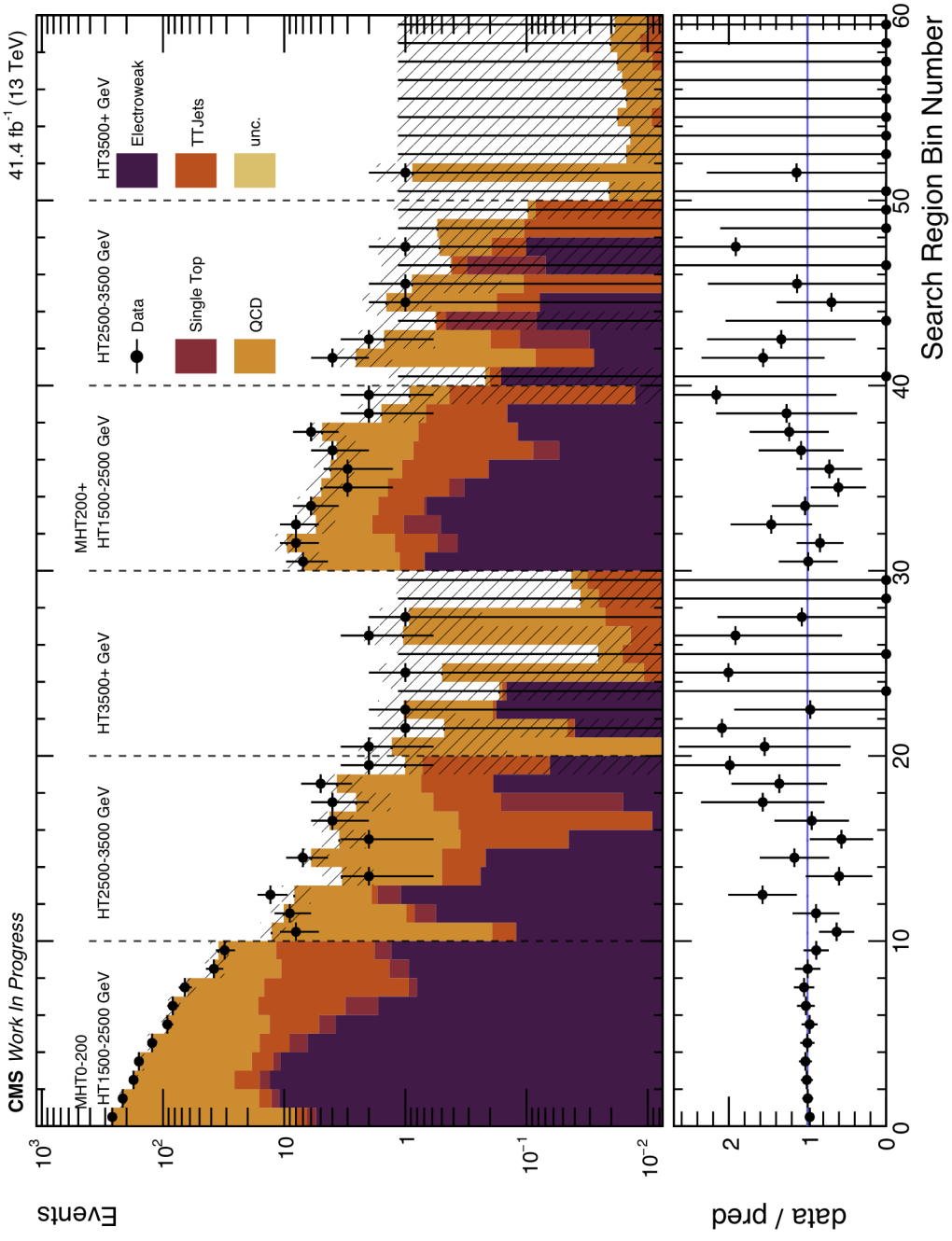


Figure 7.44: Observed and post-fit background yields across the 60 measurement bins for run year 2017.

respectively. Here we see very good agreement between background and observed event yields.

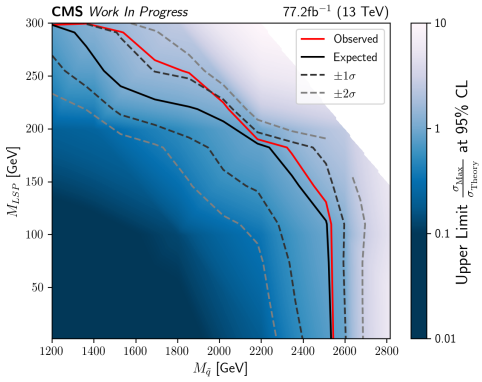
7.8.3 Signal Model Interpretations

In order to calculate upper limits on the allowed production cross-section, and thus lower bounds on $M_{\tilde{q}}$, at a confidence level of 95 %, the $\mathrm{CL_S}$ method is used, as described in Section 6.5.1.

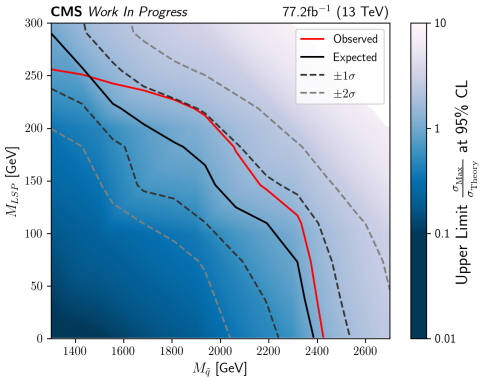
Figure 7.45 shows the observed limits at 95% CL for each of the six NMSSM mass scans, calculated for the full 77.2 fb^{-1} luminosity across run years 2016 and 2017. Also shown are the expected limits, along with $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties. Here the colour scale represents the fraction of the theoretical production cross-section which would be allowed at 95% CL, corresponding to the expected limit.

Across all mass scans we see expected lower bounds on $M_{\tilde{q}}$ between $2.2 - 2.4$ TeV in the case of the lightest 1 GeV LSP mass; considerably higher than the values of around $1 - 1.2$ TeV seen in the phenomenological work in Chapter 6 for similar LSP masses, indicating that the strategy of requiring double-b-tagged AK8 jets is effective in searching for these low-$E_{\mathrm{T}}^{\mathrm{miss}}$ NMSSM scenarios. This sensitivity remains very strong beyond $M_{\tilde{\chi}_1^0} \sim 100$ GeV, decreasing as $M_{\tilde{\chi}_1^0}$ exceeds 200 GeV, beyond which it becomes less likely for Higgs boson decays to produce sufficiently boosted $b\bar{b}$ pairs to form double-b-tagged AK8 jets.

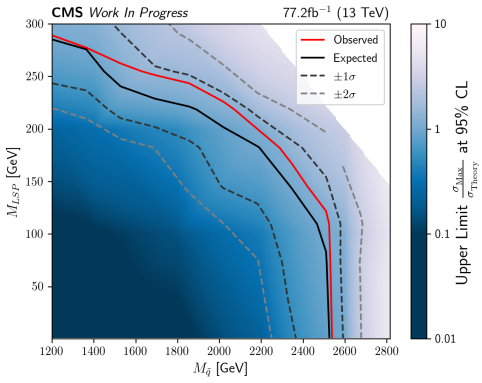
In general, up to mild fluctuations, the observed limit contours show good agreement with the expected limits, in concurrence with the agreement between background and observed yields shown in Figures 7.41 and 7.44. For higher values of $M_{\tilde{\chi}_1^0}$ we see in some cases discrepancies of the order of a 1σ deviation, with the observed limits exceeding the expected ones. These areas of parameter space correspond to bins typically in the $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV region, in which there is a slight under-estimation in the background yields, as shown in Figures 7.40 and 7.43. However we note that in the regions for which $M_{\tilde{\chi}_1^0}$ is very small, typically gen-



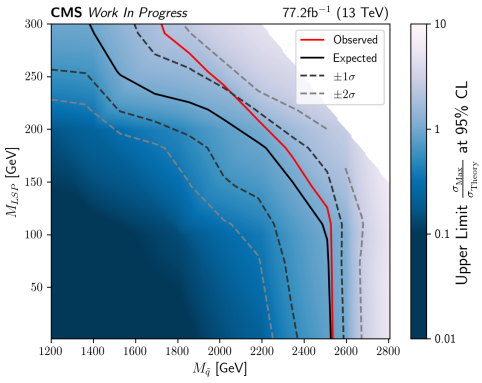
(a) BM1-type Mass Scan.



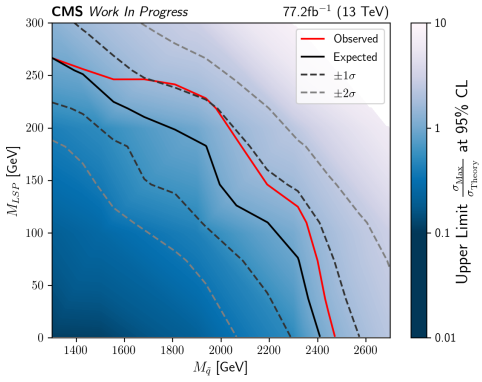
(b) BM2-type Mass Scan.



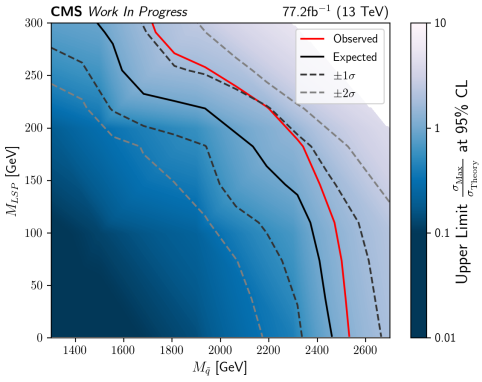
(c) BM3-type Mass Scan.



(d) BM4-type Mass Scan.



(e) BM5-type Mass Scan.



(f) BM6-type Mass Scan.

Figure 7.45: Observed and expected limits at 95% CL for the BM1–BM6-type mass scans. The X- and Y-axes represent the squark and LSP masses respectively, whilst the colour scale represents the upper limit on the strength parameter μ.

erating events in the sparsely-populated $H_{\mathrm{T}} > 3500$ GeV regions, the agreement between the observed and expected limits is very good.

We see a subtle change in behaviour in these limits when altering the mass hierarchy of the first two generations of squark and the gluino. For the BM2, BM5 and BM6 mass scans the gluino is 200 GeV lighter than the squarks, whereas for the BM1, BM3 and BM4 mass scans gluinos are only 10 GeV heavier. For a given LSP mass, mass scans BM1, BM3 and BM4 generally have expected limits around 200 GeV *stronger* than those for mass scans BM2, BM5 and BM6.

One cause of this lies in the cross-section values. As the squark and gluino masses increase, the squark-squark production cross-section increases more rapidly than that for gluino-gluino production. Indeed, beyond a squark mass of around 1.8 TeV we see the production cross-section for mass points in the BM1, BM3 and BM4 mass scans overtaken by that for points in the other mass scans featuring the same squark mass.

Furthermore as is shown in figure 2.6, BM1, BM3 and BM4 decay cascades tend to be shorter than those for BM2, BM5 and BM6, thus allowing for higher-momentum Higgs bosons, leading to a larger fraction of bottom quark jet pairs from $H \to b\bar{b}$ decays merging and being resolved as single AK8 double-b-tagged jets. As such a larger fraction of these events will be expected to pass the kinematic event selection, thus relatively strengthening the limits compared with the BM2, BM5 and BM6 mass scans.

7.8.4 Comparison With Analysis Featuring No $H_{\mathrm{T}}^{\mathrm{miss}}$ Binning

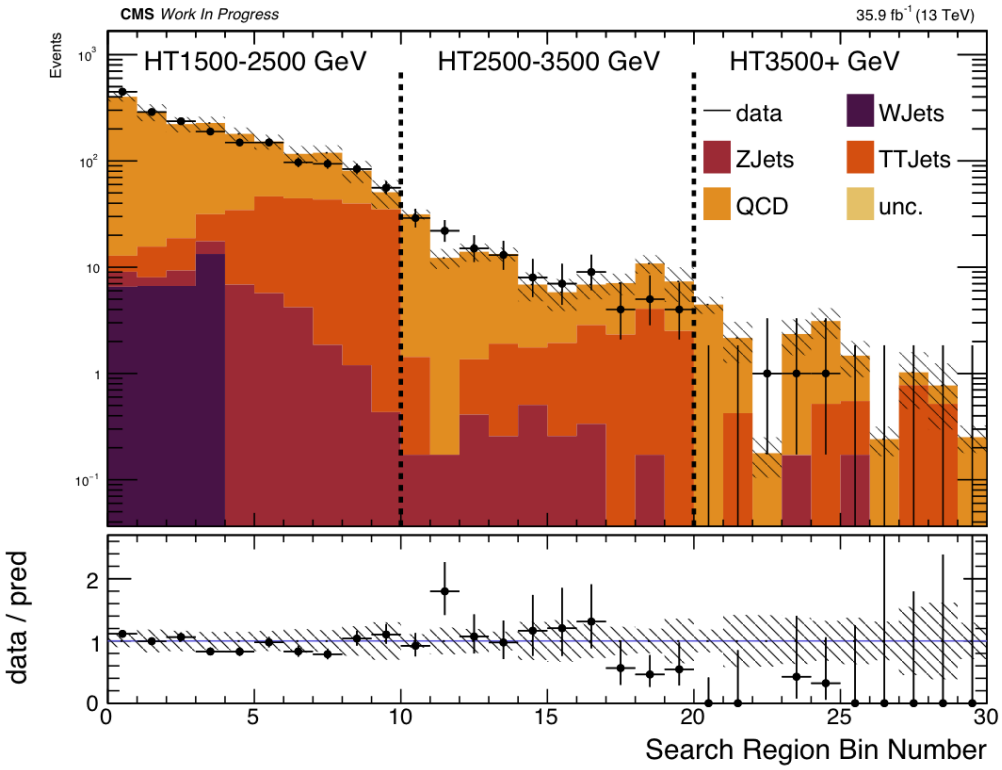
For comparison we consider the unmodified kinematic event selection as is used in Ref [2], which is the same event selection as in this analysis but featuring no $H_{\mathrm{T}}^{\mathrm{miss}}$ binning or object vetoes, in order to ascertain how much extra sensitivity is gained towards these NMSSM scenarios by the inclusion of $H_{\mathrm{T}}^{\mathrm{miss}}$ binning. Figures 7.46 and 7.47 show the distribution of expected background yields for the four combined background processes, along with the observed yields from data, for run years 2016 and 2017. The QCD yield shown is predicted using the method detailed in Section 7.5.2, with no fit applied. The observed and expected limits obtained with and without $H_{\mathrm{T}}^{\mathrm{miss}}$ binning are compared directly in Figure 7.49.

On the whole we see good agreement between the observed and predicted background yields, especially in the low-H_{T} region where we do not expect any major signal contributions, with very slight under-prediction of background yields occurs in some of the measurement bins in the mid-H_{T} region.

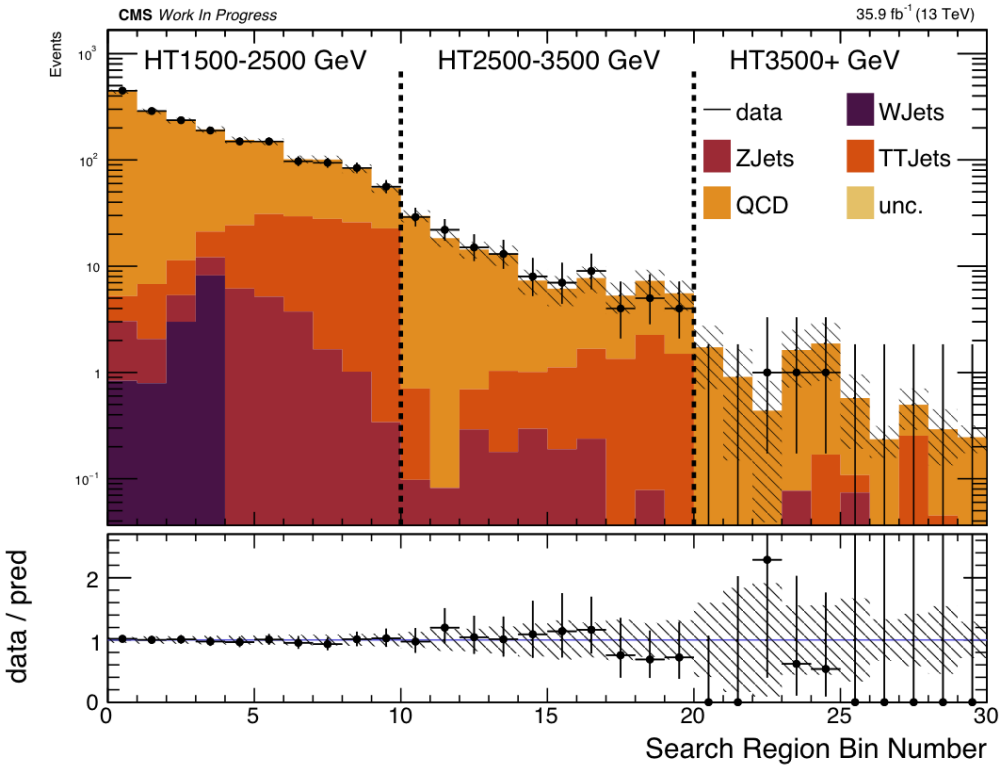
We see in Figure 7.48 a strong expected lower bound on the squark mass of around 2.4 TeV for most of these NMSSM mass scans in regions near the horizontal axis, where the LSP is very light. However, as $M_{\tilde{\chi}_1^0}$ increases this sensitivity decreases, in some cases quite quickly.

As we approach regions of parameter space where $M_{\tilde{\chi}_1^0} \sim 200$ GeV we see a sharp decrease in sensitivity, giving observed and expected lower bounds on $M_{\tilde{q}}$ around $800 - 1000$ GeV lower than is the case for $M_{\tilde{\chi}_1^0} \approx 1$ GeV, in contrast with the relatively strong limits obtained with the inclusion of $H_{\mathrm{T}}^{\mathrm{miss}}$ binning, shown in Figure 7.45. The observed limits appear a little stronger in most areas of parameter space compared with the expected limits, however remain generally within a 1σ deviation.

Figure 7.49 shows the observed and expected limits for the event selection binning in this analysis without considering $H_{\mathrm{T}}^{\mathrm{miss}}$, compared with those obtained with

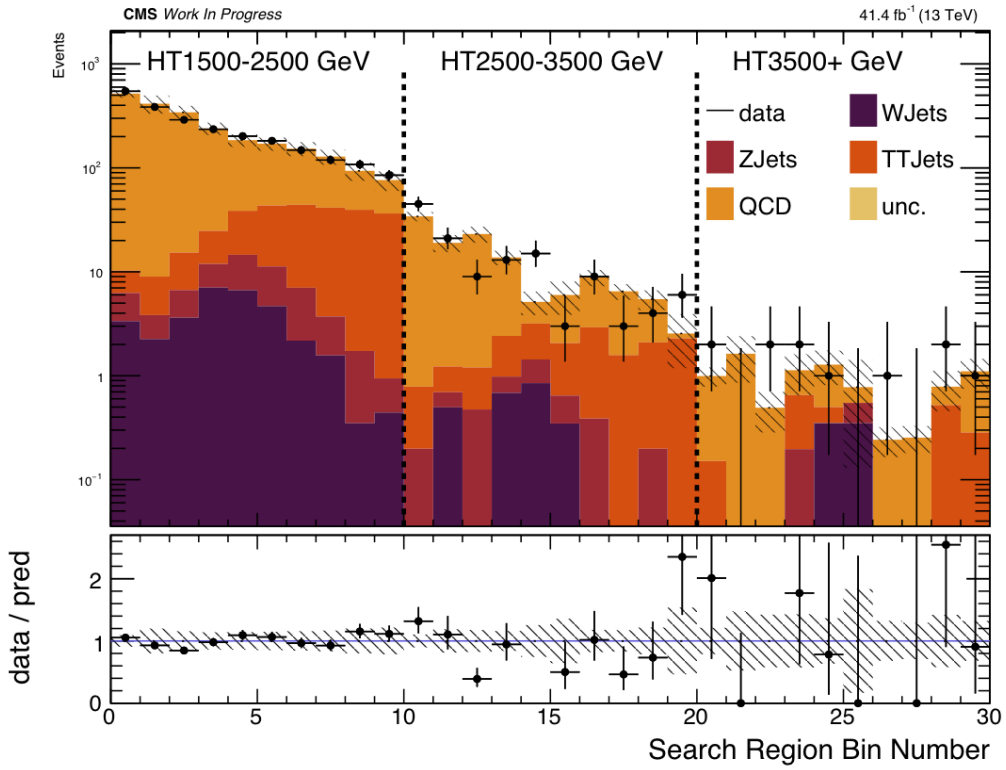


(a) Observed and pre-fit background yields.

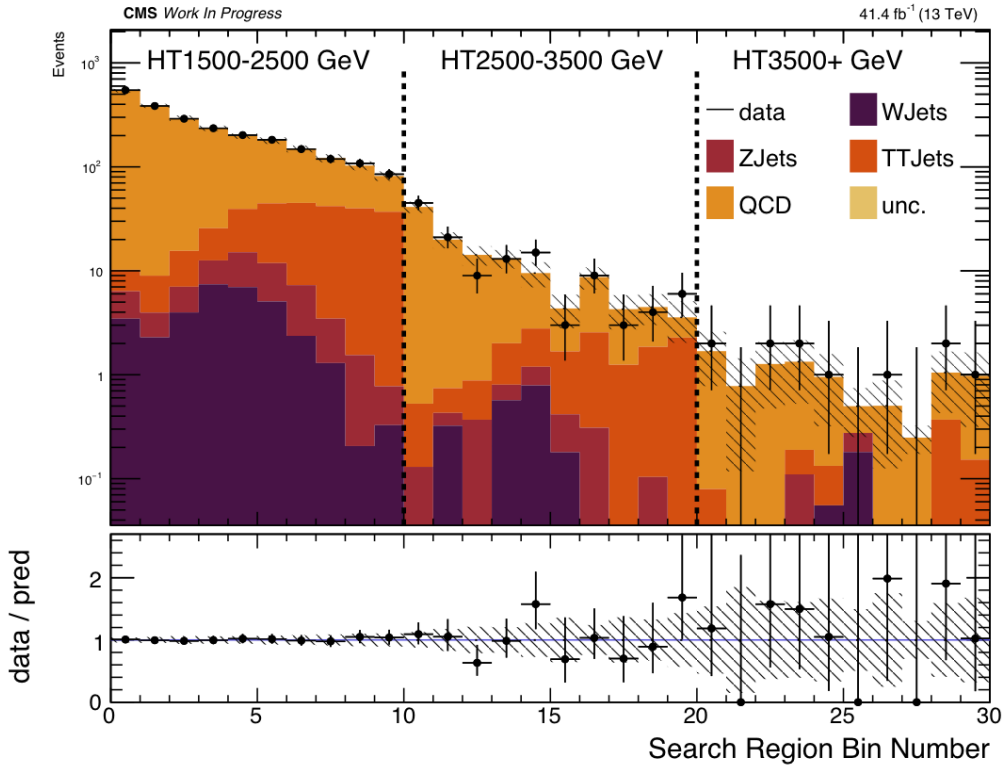


(b) Observed and post-fit background yields for the background-only hypothesis.

Figure 7.46: Observed yields compared with pre-fit and post-fit background yields in the background-only hypothesis in this analysis without $H_{\mathrm{T}}^{\mathrm{miss}}$ binning, for run-year 2016.

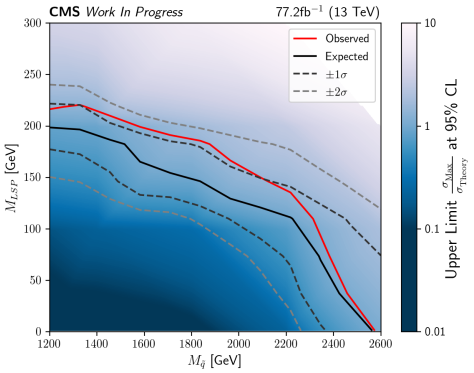


(a) Observed and pre-fit background yields.

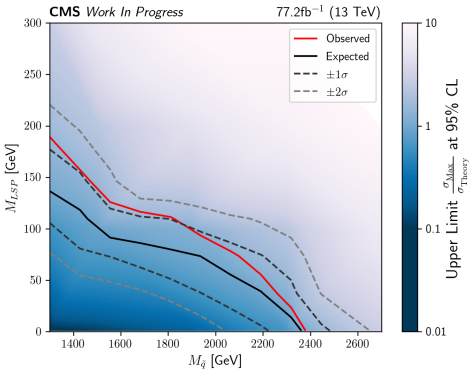


(b) Observed and post-fit background yields for the background-only hypothesis.

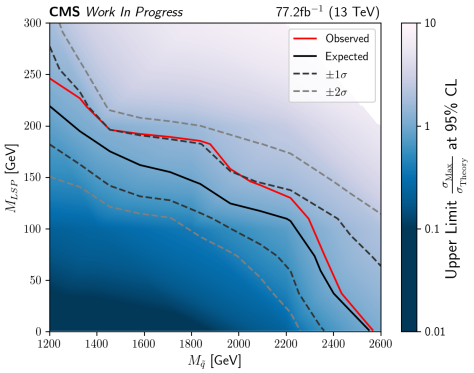
Figure 7.47: Observed yields compared with pre-fit and post-fit background yields for the background-only hypothesis in this analysis without $H_{\mathrm{T}}^{\mathrm{miss}}$ binning, for run-year 2017.



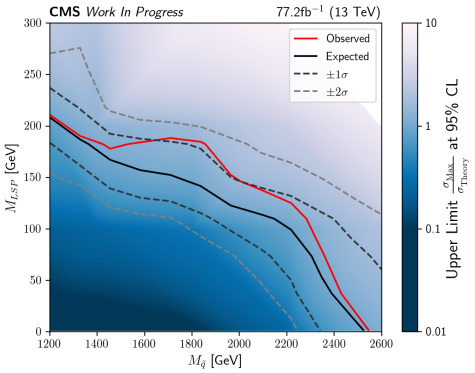
(a) BM1 Mass Scan.



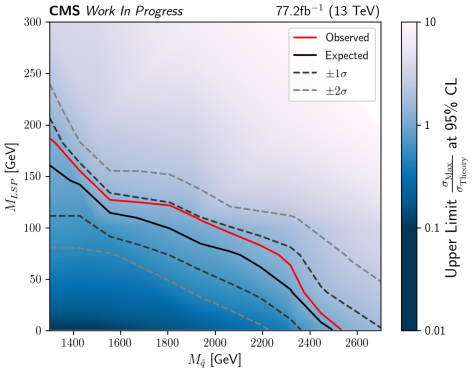
(b) BM2 Mass Scan.



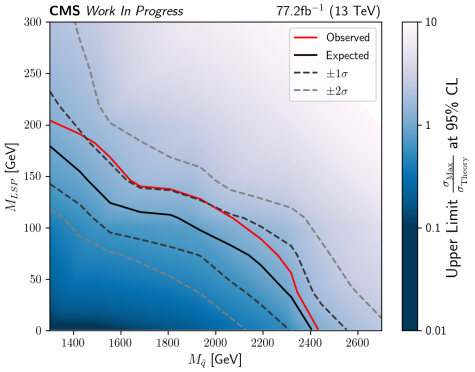
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.



(e) BM5 Mass Scan.



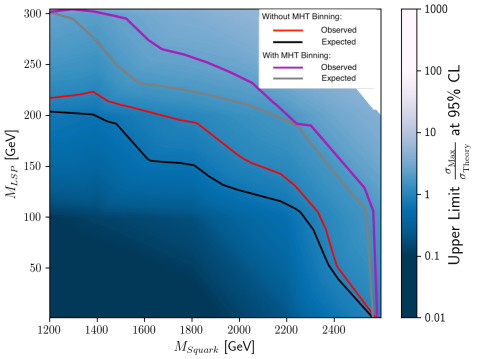
(f) BM6 Mass Scan.

Figure 7.48: Observed and expected limits at 95% CL obtained without the inclusion of $H_{\text{T}}^{\text{miss}}$ binning.

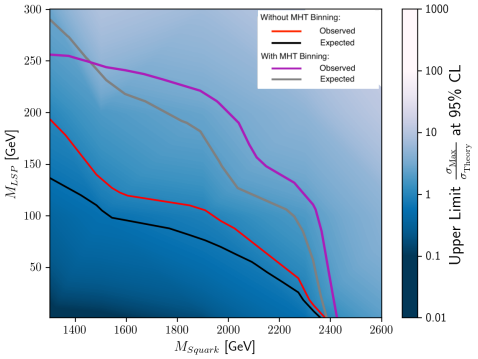
the addition of $H_{\mathrm{T}}^{\mathrm{miss}}$ bins. The observed and expected limit contours obtained without $H_{\mathrm{T}}^{\mathrm{miss}}$ binning are coloured red and black respectively, whilst the observed and expected contours obtained with the inclusion of $H_{\mathrm{T}}^{\mathrm{miss}}$ bins are coloured pink and grey.

In the case of a very light LSP we see similar values for the expected and observed limits, noting this is the area of parameter space to which the search efforts of Ref [2], without $H_{\mathrm{T}}^{\mathrm{miss}}$ binning, are exclusively tuned. However as $M_{\tilde{\chi}_1^0}$ increases as does the difference in the limits between the two analyses, with the extended binning providing as much as ~ 400 GeV extra sensitivity in $M_{\tilde{q}}$ for a mid-light LSP mass.

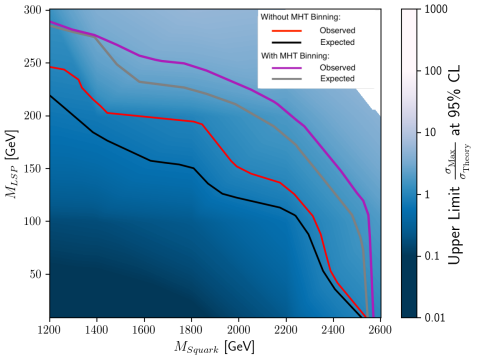
Overall we see that the sensitivity to these NMSSM scenarios has improved considerably over that seen in the phenomenological reinterpretation in Chapter 6, with the $0 < M_{\tilde{\chi}_1^0} < 300$ GeV region showing much stronger observed lower bounds on the squark mass of up to 2.4 TeV. Furthermore for the case where $M_{\tilde{\chi}_1^0}$ increases towards 300 GeV, the inclusion of measurement bins in the $H_{\mathrm{T}}^{\mathrm{miss}}$ variable allow for much higher sensitivity compared to that obtained without any $H_{\mathrm{T}}^{\mathrm{miss}}$ binning. As such, mass points within these NMSSM scenarios featuring a very light LSP and $M_{\tilde{q}} \gtrapprox 2.5$ TeV may be excluded at 95% CL, with those featuring an LSP mass of around 200 GeV being excluded in the case where $M_{\tilde{q}}$ is greater than $2 - 2.4$ TeV.



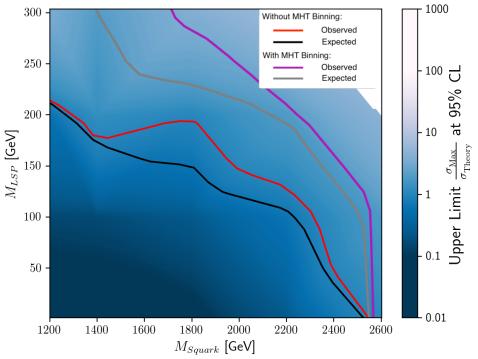
(a) BM1 Mass Scan.



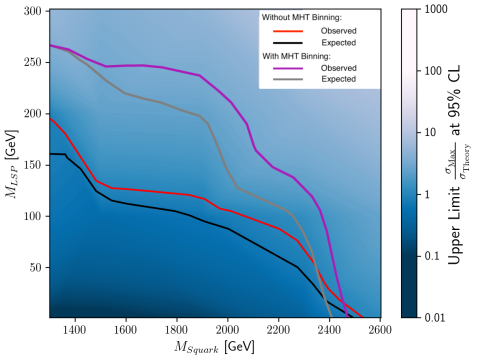
(b) BM2 Mass Scan.



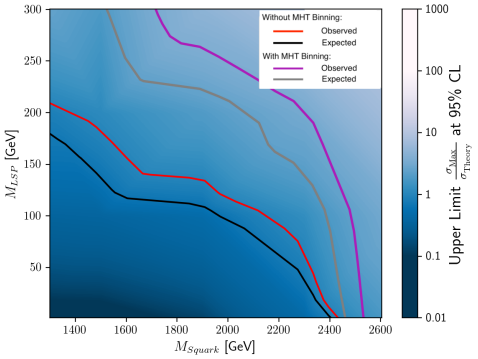
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.



(e) BM5 Mass Scan.



(f) BM6 Mass Scan.

Figure 7.49: Observed and expected limits at 95% CL obtained via this analysis, with and without binning in the $H_{\mathrm{T}}^{\mathrm{miss}}$ variable.

7.9 Summary

In this chapter we have explored how the use of a double-b-tag discriminator may be used in order to search for the NMSSM scenarios under investigation in this thesis. Considering the 2D soft-drop mass distribution of the two AK8 jets with highest double-b-tag score, signal acceptance regions were constructed around (M_H, M_H) such that the mass of a boosted Higgs boson decaying to a bottom quark-antiquark pair could be resolved.

The event yields for expected NMSSM signal and SM background processes across the 60 total H_{T}, $H_{\mathrm{T}}^{\mathrm{miss}}$ and soft-drop mass bins were then compared to those observed in data, with a likelihood model constructed, in order to derive upper bounds on the production cross-section of the respective signal models at 95% CL. This cross-section corresponds to the squark and gluino masses, and so lower bounds may then be placed on these masses, allowing for regions of parameter space to be excluded.

As seen in Figure 7.49 much larger areas of parameter space may be accessed with the inclusion of binning in the $H_{\mathrm{T}}^{\mathrm{miss}}$ variable. The lower bounds on the squark and gluino masses are essentially the same with and without this binning in the case of a very light LSP, around 2.4 TeV. However as the LSP mass increases these lower bounds decrease considerably without the inclusion of $H_{\mathrm{T}}^{\mathrm{miss}}$ binning, with the difference in the respective squark and gluino mass lower bounds with and without $H_{\mathrm{T}}^{\mathrm{miss}}$ binning being as much as 400 GeV in some cases. As we will see in Chapter 8 these results complement those obtained in 6, with the two analyses respectively probing regions for which the LSP mass is less than and greater than ~ 200 GeV.

8 | Conclusions

This thesis began by considering example benchmark models in the NMSSM whereby squark and gluino decay cascades may result in two SM-like Higgs bosons and very little $E_{\mathrm{T}}^{\mathrm{miss}}$. In order to search for such signatures, these benchmark models were developed into two-dimensional mass scans, with the squark and LSP masses characterising such scans.

The sensitivity of current major collaboration search efforts was ascertained via a phenomenological reinterpretation of the analysis in [15], which focuses on searching for SUSY in all-hadronic final states. A reduced binning scheme was considered for simplicity, considering only the measurement bins which require high H_{T}, a large number of hadronic jets, and those involving b-tagged jets.

This analysis was shown to have high sensitivity to regions of parameter space where $M_{\tilde{\chi}_1^0} > 300$ GeV, since signal events in these regions typically feature large $H_{\mathrm{T}}^{\mathrm{miss}}$, placing lower bounds on the squark and gluino masses as high as 2 TeV. However for the light LSP mass regions, especially as $M_{\tilde{\chi}_1^0} \to 0$ GeV, this sensitivity drops significantly, with the lower bound on the squark and gluino masses being as little as 1 TeV, as shown in Table 6.6.

Having found areas of parameter space to which such an analysis is not optimised, a novel analysis was carried out which focuses on the identification of high-p_{T} AK8 jets containing a boosted Higgs boson decaying to a bottom quark-antiquark pair. Upon finding the baseline selection appropriate for the purposes of studying the NMSSM scenarios under investigation in this thesis, the analysis was extended in order to better probe the regions of parameter space for which the LSP is light, but more than just a few GeV.

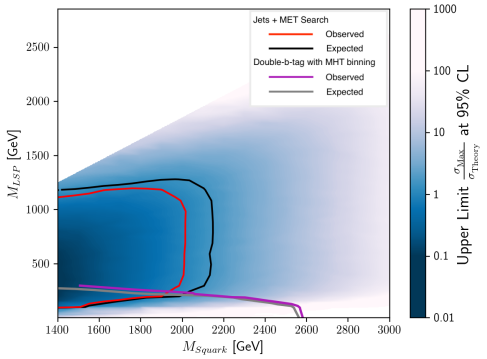
The introduction of binning in the $H_{\mathrm{T}}^{\mathrm{miss}}$ variable allowed for retaining the sensitivity to the light-LSP scenarios, most of which events fall into the $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV category, whilst also providing sensitivity to mass points featuring LSP masses across the $[0, 300)$ GeV range. As can be seen in Figure 7.49 the sensitivity to

mass points with a $M_{\tilde{\chi}_1^0} \sim 1$ GeV is essentially the same as that obtained without $H_{\mathrm{T}}^{\mathrm{miss}}$ binning, whereas the inclusion of such binning allows for the exclusion of a much larger area of parameter space for these NMSSM models.

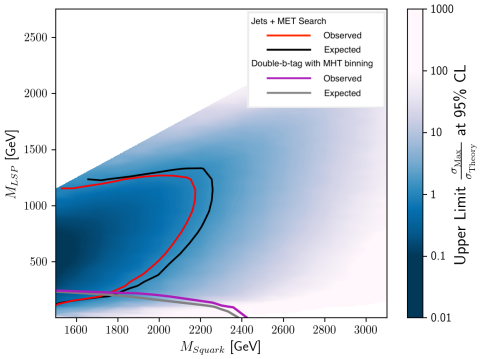
Alas, as shown in Figures 7.41 and 7.44, no significant excesses have been observed in data as part of the experimental search effort in this thesis, with the observed limit contours in Section 7.8.3 deviating, accordingly, very little from the expected limits. Figure 8.1 compares the lower bounds placed on the squark and gluino masses for the phenomenological work in Chapter 6 and the experimental analysis work in Chapter 7 respectively, across the two-dimensional mass scans considered in this thesis. The observed and expected limits from Chapter 6 are respectively coloured red and black, whilst the observed and expected limits from Chapter 7 are respectively coloured pink and grey.

In the case of a very light LSP the analysis in Chapter 7 allowed for the probing of a much larger region of phase space in the NMSSM scenarios investigated in this thesis, with the lower bounds placed on the squark and gluino masses increasing from around 1 TeV [1], shown in the phenomenological reinterpretation of [15], which was shown in Chapter 6, to as much as $\sim$ 2.5 TeV. The observed and expected limit contours typically meet at around $M_{\tilde{\chi}_1^0} \approx 200$ GeV, with the lower bound on the squark mass in the range of $1.8 - -2$ TeV, allowing for a large amount of parameter space to be excluded at 95% CL for the NMSSM scenarios considered in this thesis.

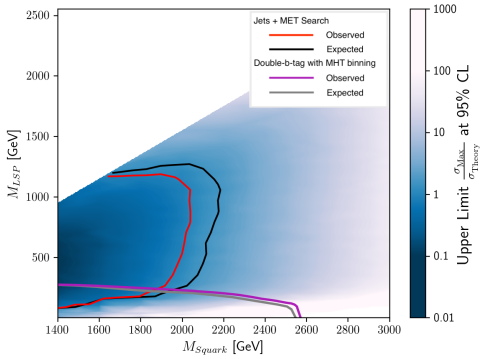
Similarly, no ATLAS or CMS collaboration search has thus far produced any evidence for BSM physics. Confidence in the elegant solution that is SUSY may be waning over time, with lower bounds on squark masses pushing the allowed parameter space further from the domain in which SUSY provides a solution to the hierarchy problem. However, the NMSSM scenarios in this thesis provide just one example of how the parameters of a model may allow for a stealthy manifestation in nature, providing further search challenges at the LHC and future



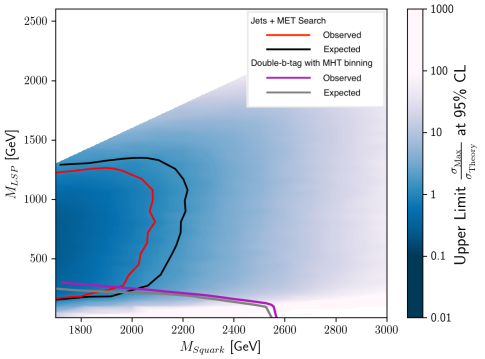
(a) BM1 Mass Scan.



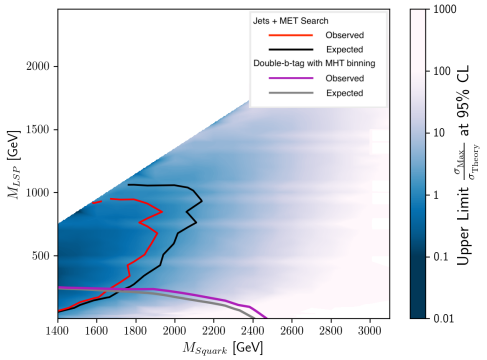
(b) BM2 Mass Scan.



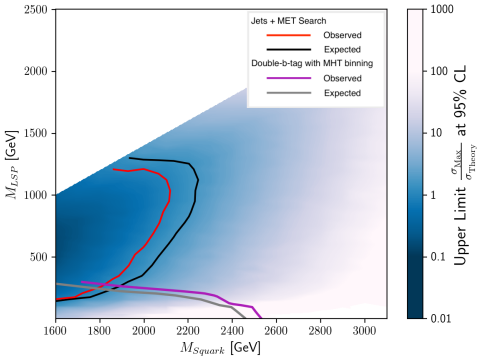
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.



(e) BM5 Mass Scan.



(f) BM6 Mass Scan.

Figure 8.1: Observed (pink) and expected (grey) limits at 95% CL for the experimental analysis in Chapter 7, compared with the observed (red) and expected (black) limits obtained via the phenomenological reinterpretation of [15] in Chapter 6.

collider experiments. Furthermore R-parity-violating SUSY models, whilst not providing a natural DM candidate, may escape $E_{\mathrm{T}}^{\mathrm{miss}}$-based searches.

Whilst the specific models under consideration in this thesis have been placed under pressure by searches such as [15] and the analysis in Chapter 7, it is clear that the SM does not form a comprehensive model of physics up to the GUT scale. DM searches are ongoing at direct detection experiments as well as collider experiments, providing more simplified models for which to search.

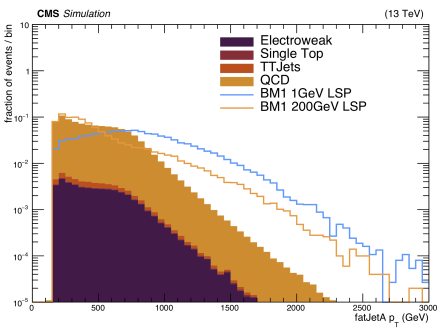
However, the continually increasing collection of search results casts ever more doubt onto the prospects of finding SUSY at the LHC. Whilst possibilities such as the unification of the strong and electroweak couplings at the GUT scale [13] continue to push in favour of SUSY, further driving the enormous search effort at experiments at the LHC, it remains to be seen whether SUSY will turn out to be anything more than a beautiful collection of mathematical work.

The current upgrade work taking place at the LHC experiments as well as plans for the High Luminosity LHC (HL-LHC) [110] and the potential Future Circular Collider (FCC) [111] indicate the next few decades will be pivotal in the search for BSM physics. A consequence of the advances in collision rate and centre-of-mass energy is of course enormous challenges in terms of data acquisition and analysis. As such the near future promises to be an exciting time in high-energy particle physics, and will give the best chance yet of finding evidence for new particles not described by the Standard Model.

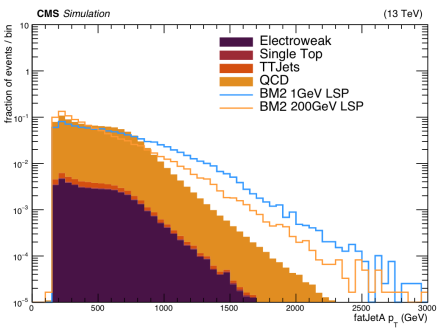
Appendix A | Event Selection: Remaining Mass Points

A.1 AK4 and AK8 jet p_{T}

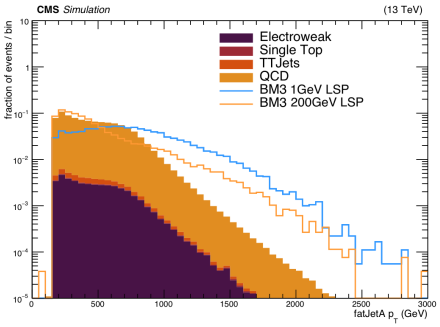
Figures A.1 and A.2 show the normalised AK8 and AK4 jet p_{T} distributions for example NMSSM signal mass points compared with those for SM background processes.



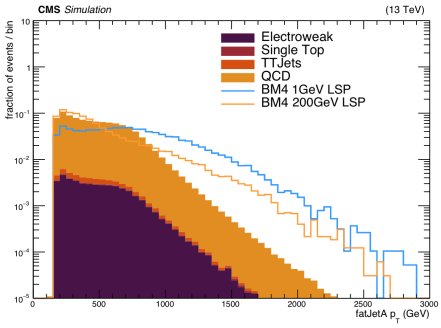
(a) BM1 Mass Scan.



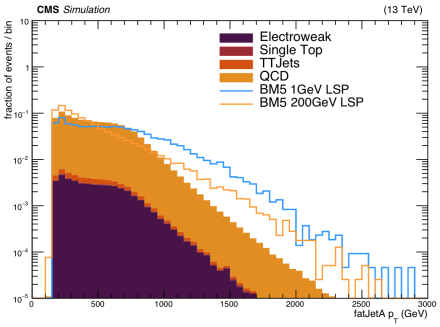
(b) BM2 Mass Scan.



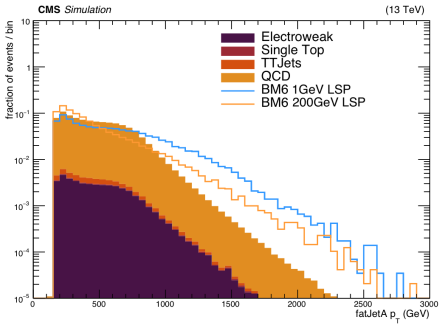
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.

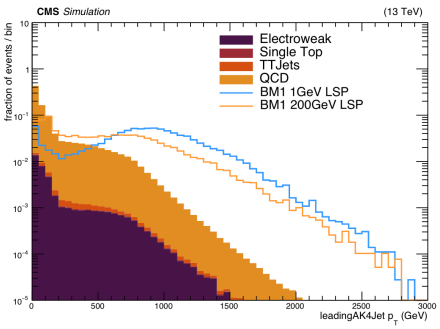


(e) BM5 Mass Scan.

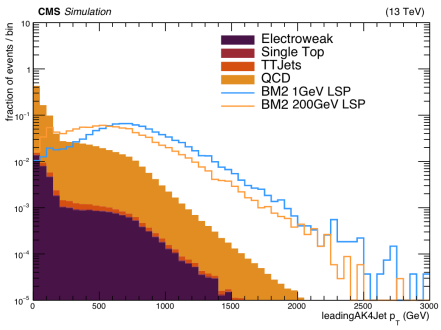


(f) BM6 Mass Scan.

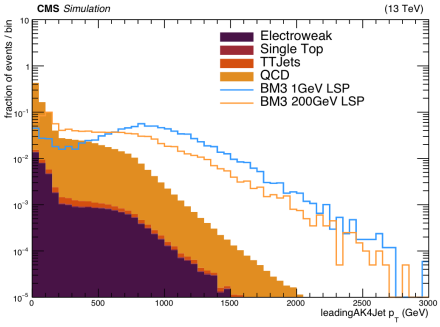
Figure A.1: AK8 jet p_{T} comparison between example mass points and SM background processes.



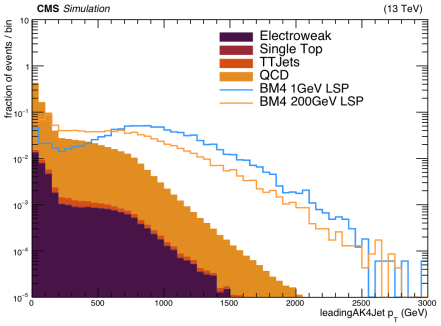
(a) BM1 Mass Scan.



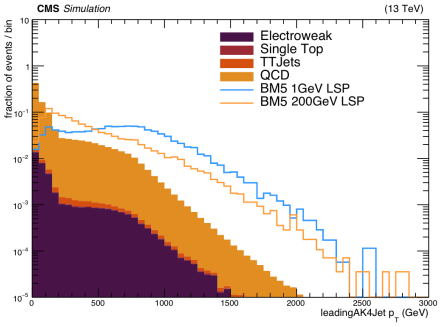
(b) BM2 Mass Scan.



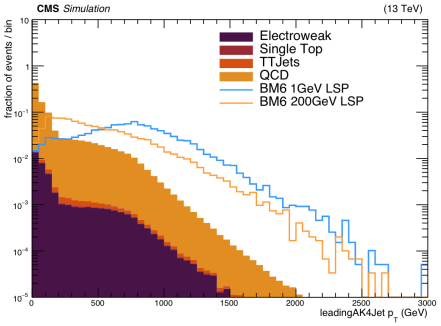
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.



(e) BM5 Mass Scan.



(f) BM6 Mass Scan.

Figure A.2: Leading AK4 jet p_{T} comparison between example mass points and SM background processes.

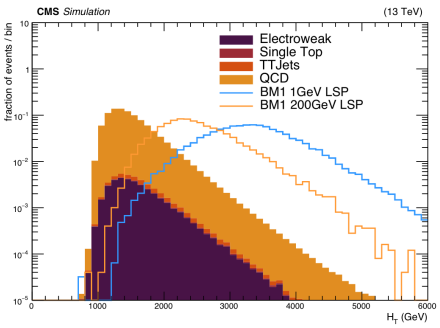
A.2 H_{T} binning

As the squark, gluino and LSP masses vary, so does the mean H_{T} value. As is shown in Figure 7.2, varying the NLSP and LSP masses has a notable effect on the H_{T} distribution, since a heavier LSP inherits a larger fraction of the NLSP momentum.

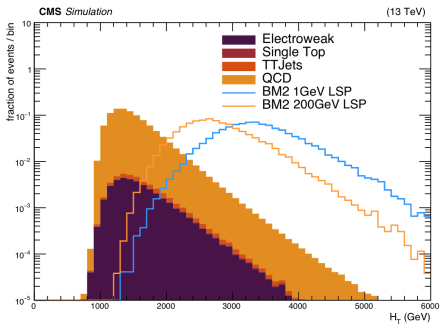
Therefore, in order to improve discrimination between signal+background and background-only expected yields, the H_{T} measurement is split into three bins:

$$H_{\mathrm{T}} \in [1500, 2500), [2500, 3500), [3500, \infty) \text{ GeV}$$

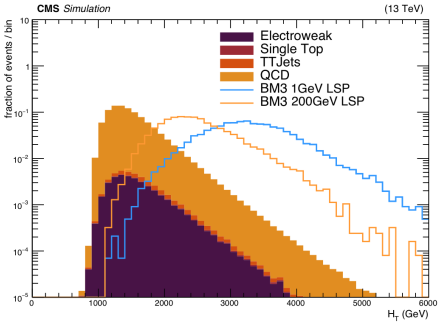
As is also shown in Figure 7.2 the background process yields will be reduced considerably following this H_{T} requirement. However, owing to the large production cross-section values for processes such as QCD and $t\bar{t}$ further background reduction steps are required.



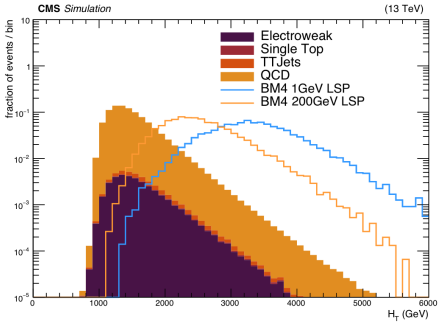
(a) BM1 Mass Scan.



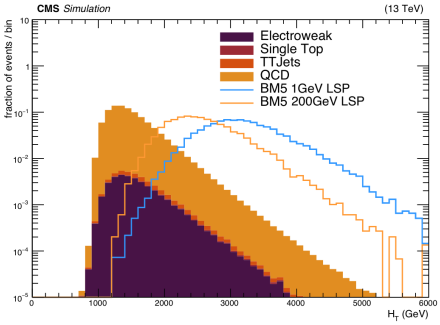
(b) BM2 Mass Scan.



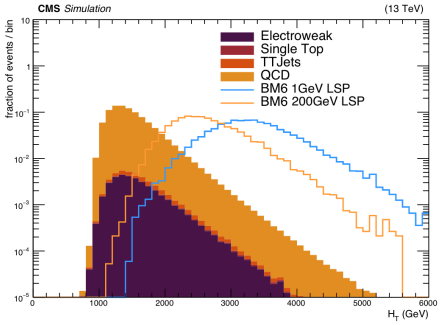
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.



(e) BM5 Mass Scan.



(f) BM6 Mass Scan.

Figure A.3: Scalar H_T comparison between example mass points and SM background processes.

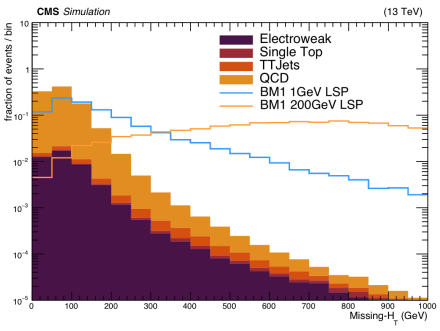
A.3 $H_{\mathrm{T}}^{\mathrm{miss}}$ binning

Whilst a very light LSP will highly suppress any observed $H_{\mathrm{T}}^{\mathrm{miss}}$, increasing this mass quickly generates a larger $H_{\mathrm{T}}^{\mathrm{miss}}$ signal, especially in the case where the LSP is heavier than the Higgs boson, thus inheriting the lion's share of the NLSP momentum. This is illustrated in Figure 7.12.

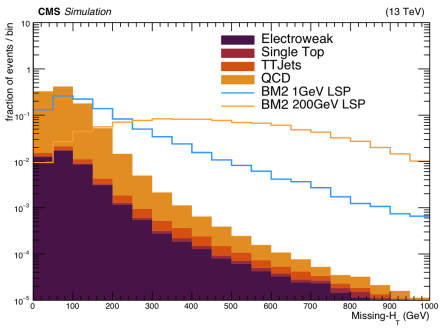
Therefore, in order to better probe the breadth of parameter space for which $0 < M_{\mathrm{LSP}} < 200$ GeV, we split the $H_{\mathrm{T}}^{\mathrm{miss}}$ range into two bins:

$$H_{\mathrm{T}}^{\mathrm{miss}} \in [0, 200), [200, \infty) \text{ GeV}.$$

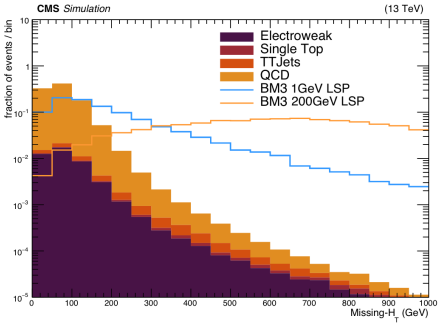
It may be seen in Figure 7.12 that for mass points with very light LSP the majority of events will fall into the lower $H_{\mathrm{T}}^{\mathrm{miss}}$ bin. However, since this will also be the case for the main background processes we expect a smaller number of background events in bins for which $H_{\mathrm{T}}^{\mathrm{miss}} \in [200, \infty)$ GeV. As a result, sensitivity to scenarios with heavier LSP masses should benefit considerably.



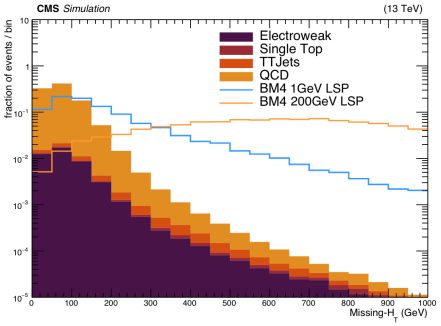
(a) BM1 Mass Scan.



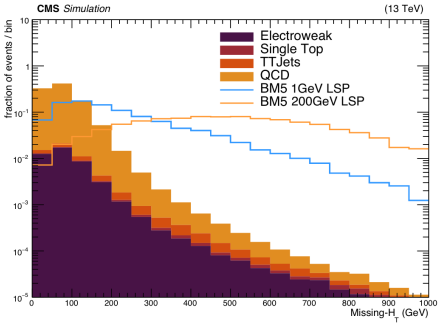
(b) BM2 Mass Scan.



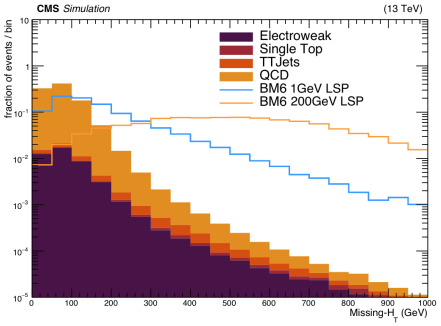
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.



(e) BM5 Mass Scan.



(f) BM6 Mass Scan.

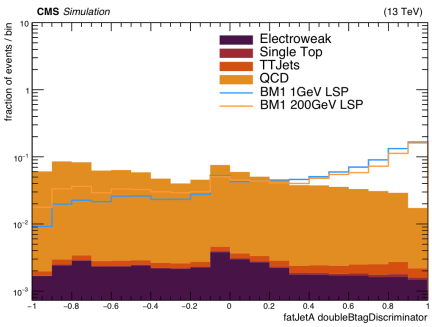
Figure A.4: Missing-H_{T} comparison between example mass points and SM background processes.

A.4 2D AK8 Double-b-tag Score Selection

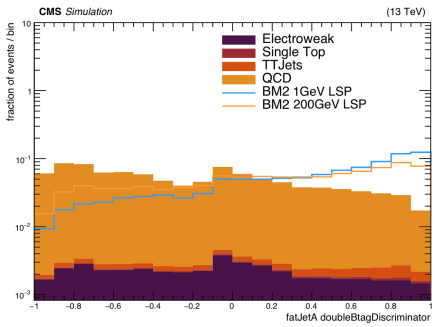
Since the final state under consideration is that for which two Higgs bosons each decay into a boosted bottom quark-antiquark pair, too closely overlapping to allow for resolving the quarks as individual b-tagged jets, the focus instead is to construct a larger-radius jet cone around each $b\bar{b}$ pair. This jet, having twice the radius parameter of a regular AK4 jet, is then assigned a discriminator score indicating whether it has been identified as containing two bottom quarks stemming from the same object – in this case, a Higgs boson.

This discriminator score is driven by a Boosted Decision Tree (BDT), which uses 27 measurement variables to determine the output, between -1 and 1, representing the likelihood for a given jet to have originated from an object decaying into a bottom quark-antiquark pair.

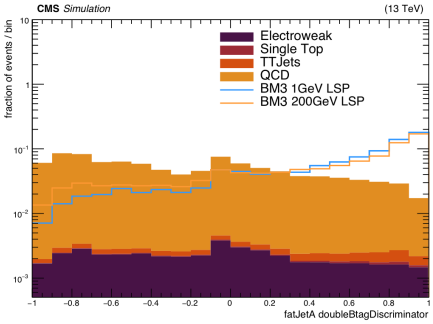
Consider the two central AK8 jets, each having $p_T > 300$ GeV, with the highest double-b-tag discriminator score. We denote these jets A and B, in random order, with double-b-tag scores D_{bb}^{A} and D_{bb}^{B} respectively. Figure 7.5 shows the distribution of the double-b-tag score of AK8 jet A for example signal mass points compared with SM background processes. It is clear here that these NMSSM scenarios generate events containing AK8 jets with very high double-b-tag scores, owing to the highly boosted Higgs boson decays, whilst background processes generally give a more flat distribution.



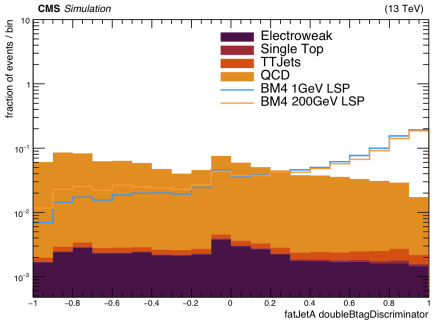
(a) BM1 Mass Scan.



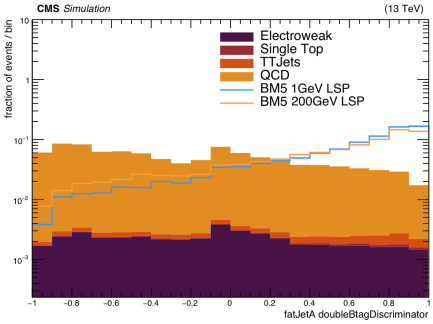
(b) BM2 Mass Scan.



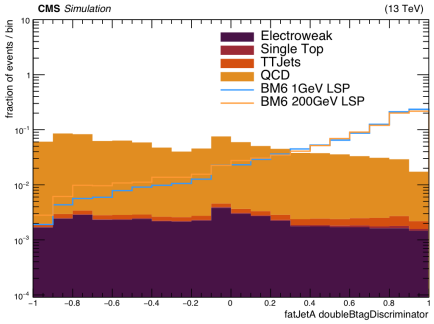
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.



(e) BM5 Mass Scan.

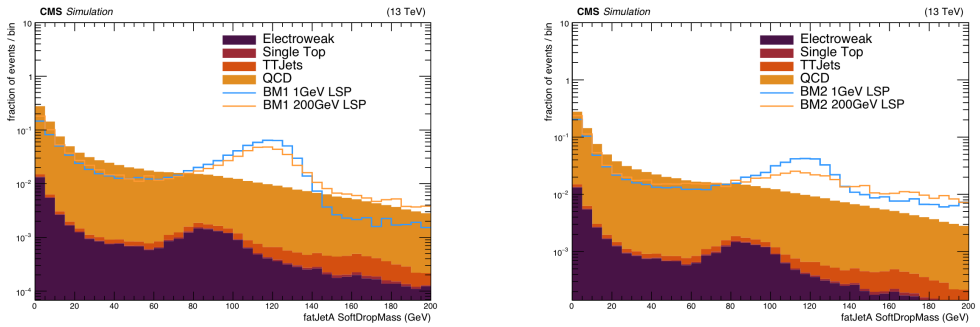


(f) BM6 Mass Scan.

Figure A.5: Double-*b*-tag score comparison between example mass points and SM background processes.

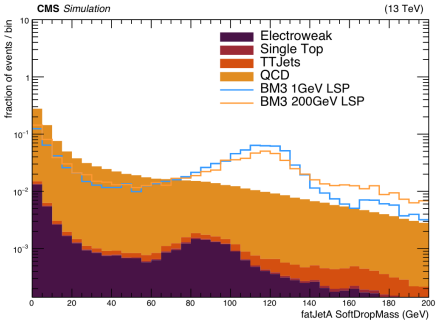
A.5 2D AK8 soft-drop Mass Binning

Figure A.6 shows a comparison between the soft-drop mass for the AK8 jet with highest double-b-tag score for example mass points in the NMSSM scenarios under investigation in this thesis, with that for SM background processes. We see a notable bump around 125 GeV, corresponding to the SM-like Higgs boson mass for the signal processes. The QCD background generally features a decreasing trend towards higher soft-drop mass values, with no bumps, whilst the electroweak background features small bumps around the W and Z boson masses.

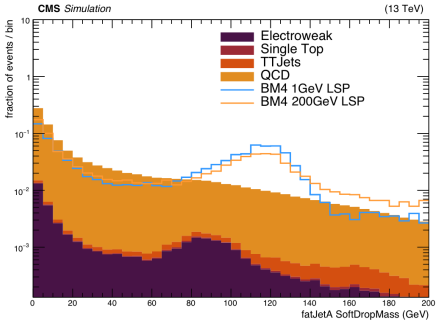


(a) BM1 Mass Scan.

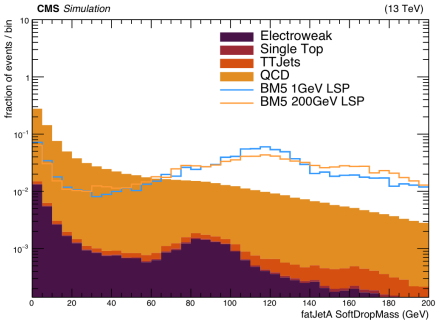
(b) BM2 Mass Scan.



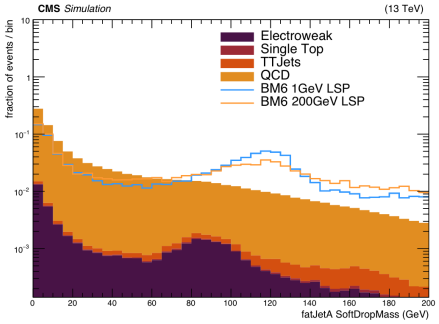
(c) BM3 Mass Scan.



(d) BM4 Mass Scan.



(e) BM5 Mass Scan.



(f) BM6 Mass Scan.

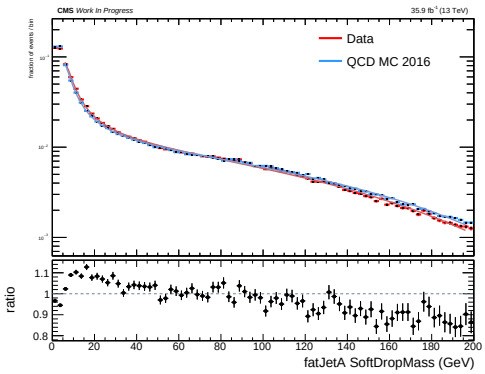
Figure A.6: AK8 jet soft-drop mass comparison between example mass points and SM background processes.

Appendix B | Data-Driven QCD Estimation

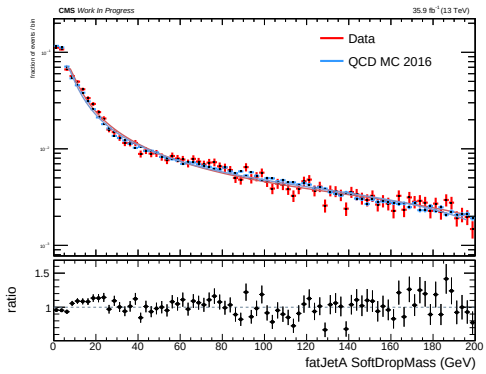
B.1 QCD AK8 soft-drop Mass Fit Function

Figure B.1 below shows the 1D AK8 soft-drop mass distributions along with polynomial fit for data and QCD MC in the QCD-dominated anti-tag double-b-tag score region. The twelve plots represent the six H_{T}-$H_{\mathrm{T}}^{\mathrm{miss}}$ bins across run years 2016 and 2017.

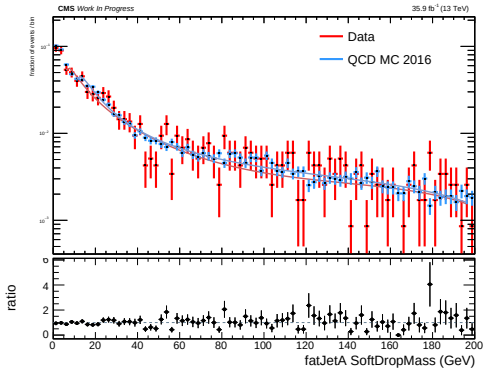
Low statistics may be seen for data in both $H_{\mathrm{T}}^{\mathrm{miss}}$ regions within the highest H_{T} region, as well as the $2500 < H_{\mathrm{T}} < 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV region. As such in these regions QCD MC simulated events are used to derive the F_i factors instead of data.



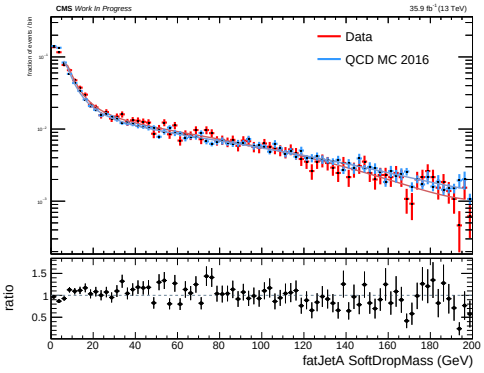
(a) $1500 < H_{\mathrm{T}} < 2500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2016.



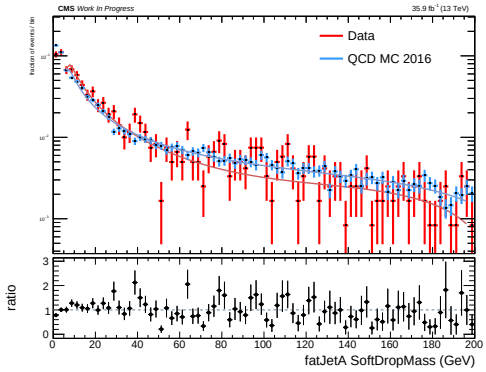
(b) $2500 < H_{\mathrm{T}} < 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2016.



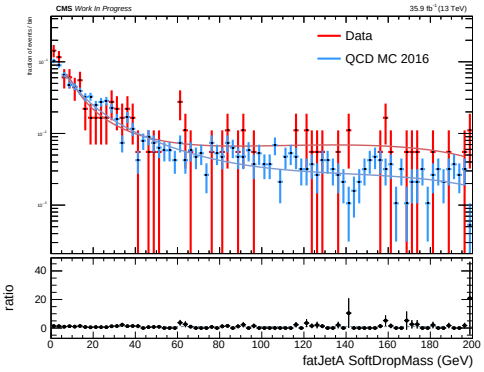
(c) $H_{\mathrm{T}} > 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2016.



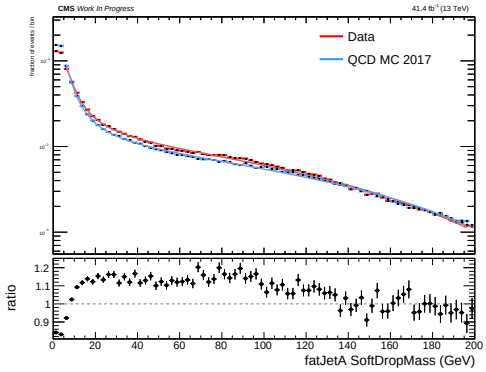
(d) $1500 < H_{\mathrm{T}} < 2500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV for run year 2016.



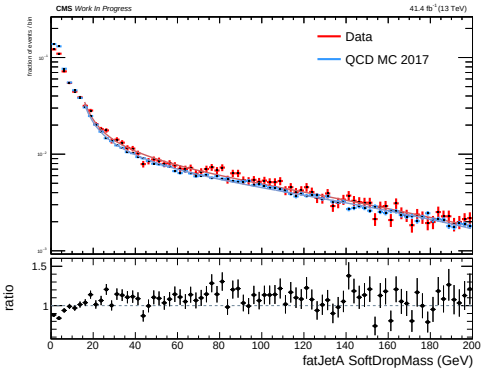
(e) $2500 < H_{\mathrm{T}} < 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV for run year 2016.



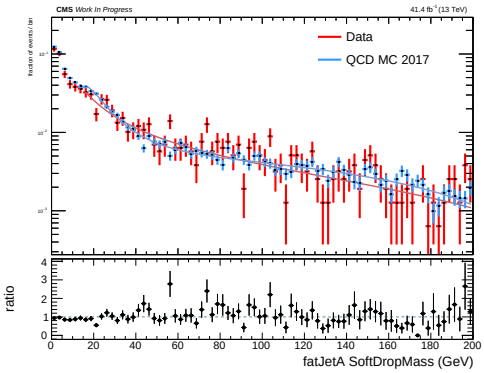
(f) $H_{\mathrm{T}} > 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV for run year 2016.



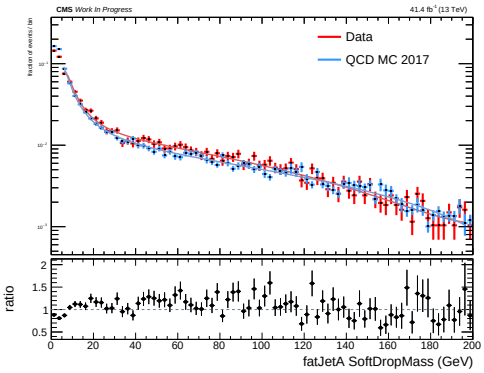
(g) $1500 < H_{\mathrm{T}} < 2500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2017.



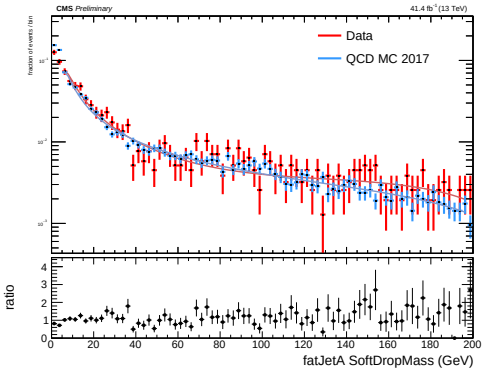
(h) $2500 < H_{\mathrm{T}} < 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2017.



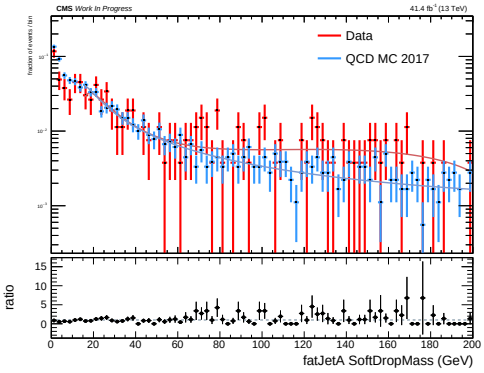
(i) $H_{\mathrm{T}} > 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV for run year 2017.



(j) $1500 < H_{\mathrm{T}} < 2500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV for run year 2017.



(k) $2500 < H_{\mathrm{T}} < 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV for run year 2017.



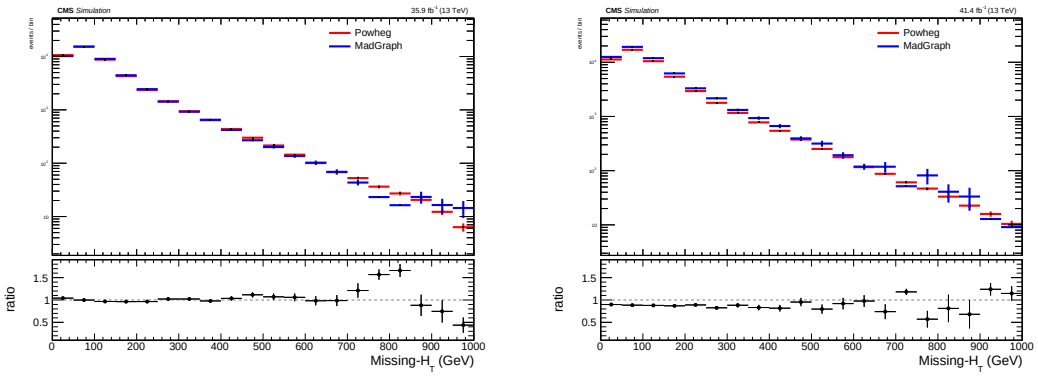
(l) $H_{\mathrm{T}} > 3500$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV for run year 2017.

Figure B.1: Rows 1–3: The normalised soft-drop mass distributions of fatJetA, along with fits, for the six H_{T}-$H_{\mathrm{T}}^{\mathrm{miss}}$ regions in the anti-tag double-b-tag region for data and QCD MC for run year 2016. Rows 5–6: As above for run year 2017.

Appendix C | Further Comparison Between $t\bar{t}$ Samples

Figures C.1 to C.6 show direct comparisons between the Powheg NLO and MadGraph LO H_{T}-binned $t\bar{t}$ samples for a wide selection of measurement variables. With the exception of $H_{\mathrm{T}}^{\mathrm{LHE}}$, excellent agreement is observed between these samples; thus further validating the use of the high-H_{T} MadGraph samples in order to improve statistics in the extreme H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$ regions considered in this analysis.

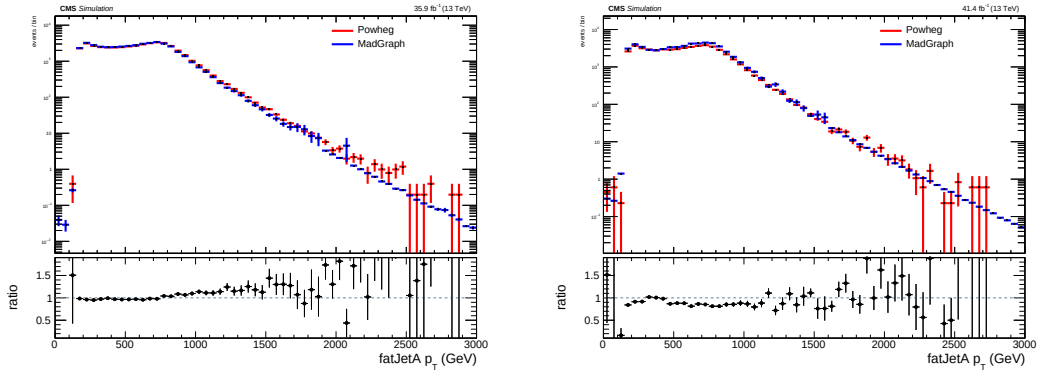
In all cases a basic pre-selection is imposed, considering only events which contain at least two AK8 jets, with implicit $p_{\mathrm{T}}^{\mathrm{AK8}} > 170$ GeV requirement, and have $H_{\mathrm{T}} > 1500$ GeV.



(a) Run year 2016.

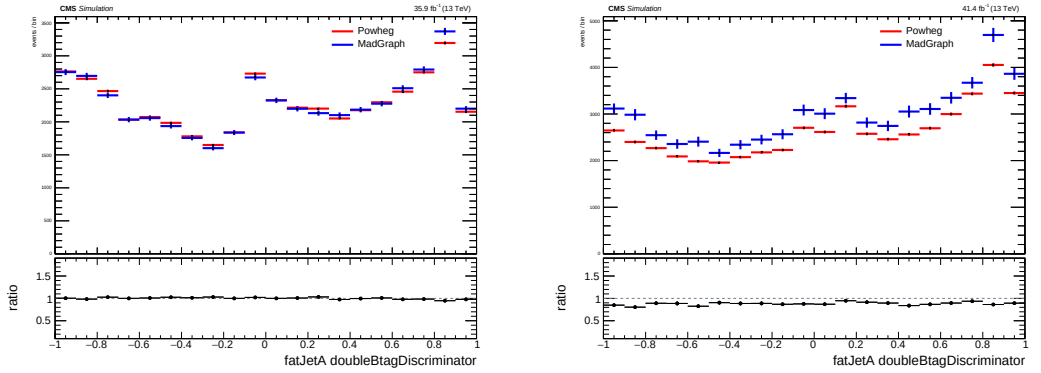
(b) Run year 2017.

Figure C.1: $H_{\mathrm{T}}^{\mathrm{miss}}$ comparison between Powheg NLO inclusive and MadGraph LO H_{T}-binned $t\bar{t}$ MC samples.



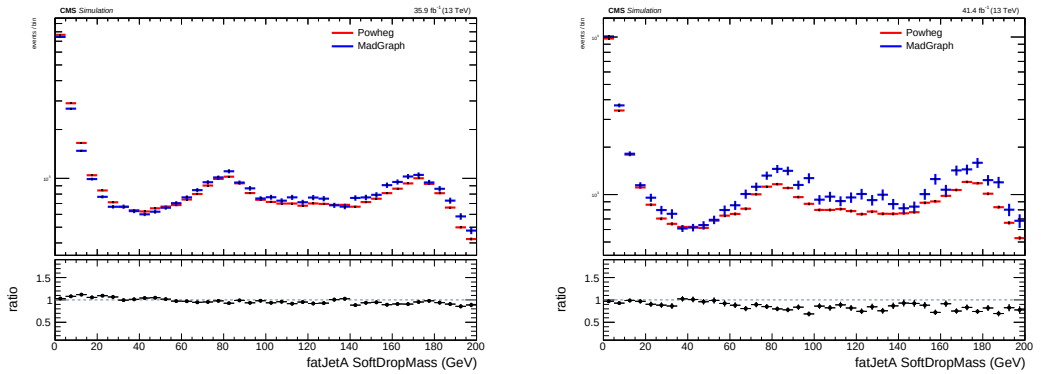
(a) Run year 2016. (b) Run year 2017.

Figure C.2: AK8 jet p_{T} comparison between Powheg NLO inclusive and MadGraph LO H_{T}-binned $t\bar{t}$ MC samples.



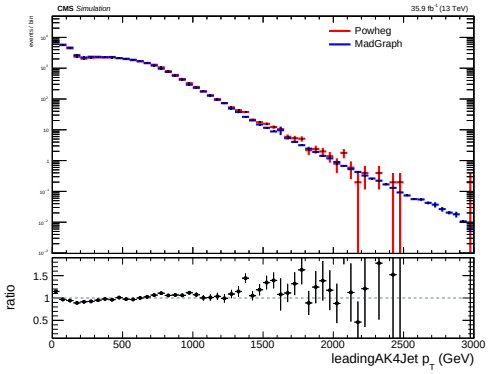
(a) Run year 2016. (b) Run year 2017.

Figure C.3: AK8 jet double-b-tag score comparison between Powheg NLO inclusive and MadGraph LO H_{T}-binned $t\bar{t}$ MC samples.

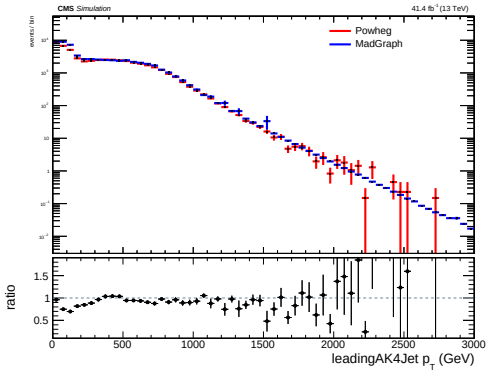


(a) Run year 2016. (b) Run year 2017.

Figure C.4: AK8 jet soft-drop mass comparison between Powheg NLO inclusive and MadGraph LO H_{T}-binned $t\bar{t}$ MC samples.

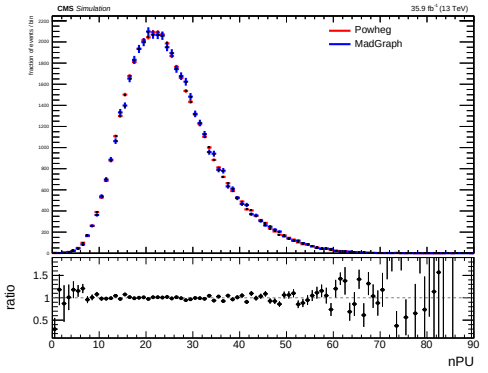


(a) Run year 2016.

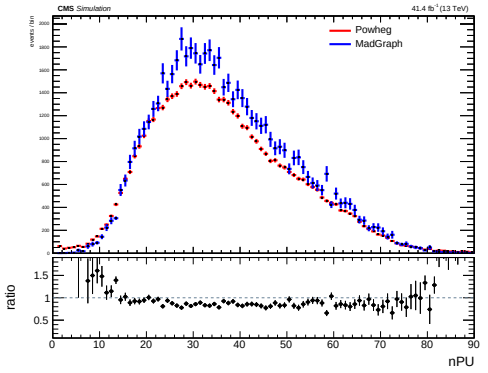


(b) Run year 2017.

Figure C.5: Leading AK4 jet p_{T} comparison between Powheg NLO inclusive and MadGraph LO H_{T}-binned $t\bar{t}$ MC samples.



(a) Run year 2016.



(b) Run year 2017.

Figure C.6: n_{PU} comparison between Powheg NLO inclusive and MadGraph LO H_{T}-binned $t\bar{t}$ MC samples.

Appendix D | Event Yields for Remaining Mass Scans

As is shown in Figures 7.39 and 7.42 for the BM1 mass scan, Figures D.1 to D.5 show the distribution of expected background yields for the four combined background processes, along with the predicted yields for three example signal mass points from each of the BM2–6 mass scans respectively.

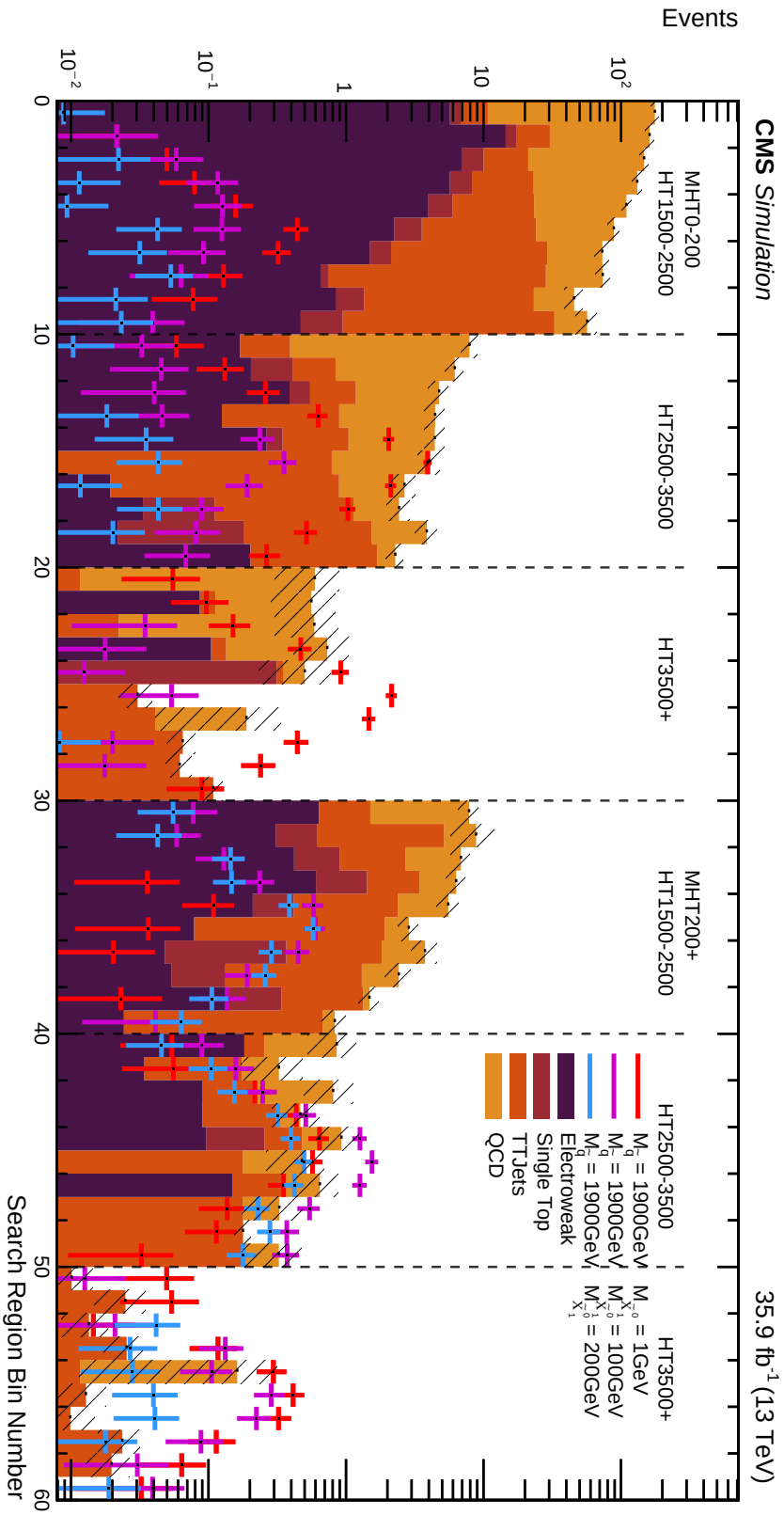


Figure D.1: Expected yields across the sixty measurement bins for three example signal mass points from the BM2 mass scan compared with SM MC background processes.

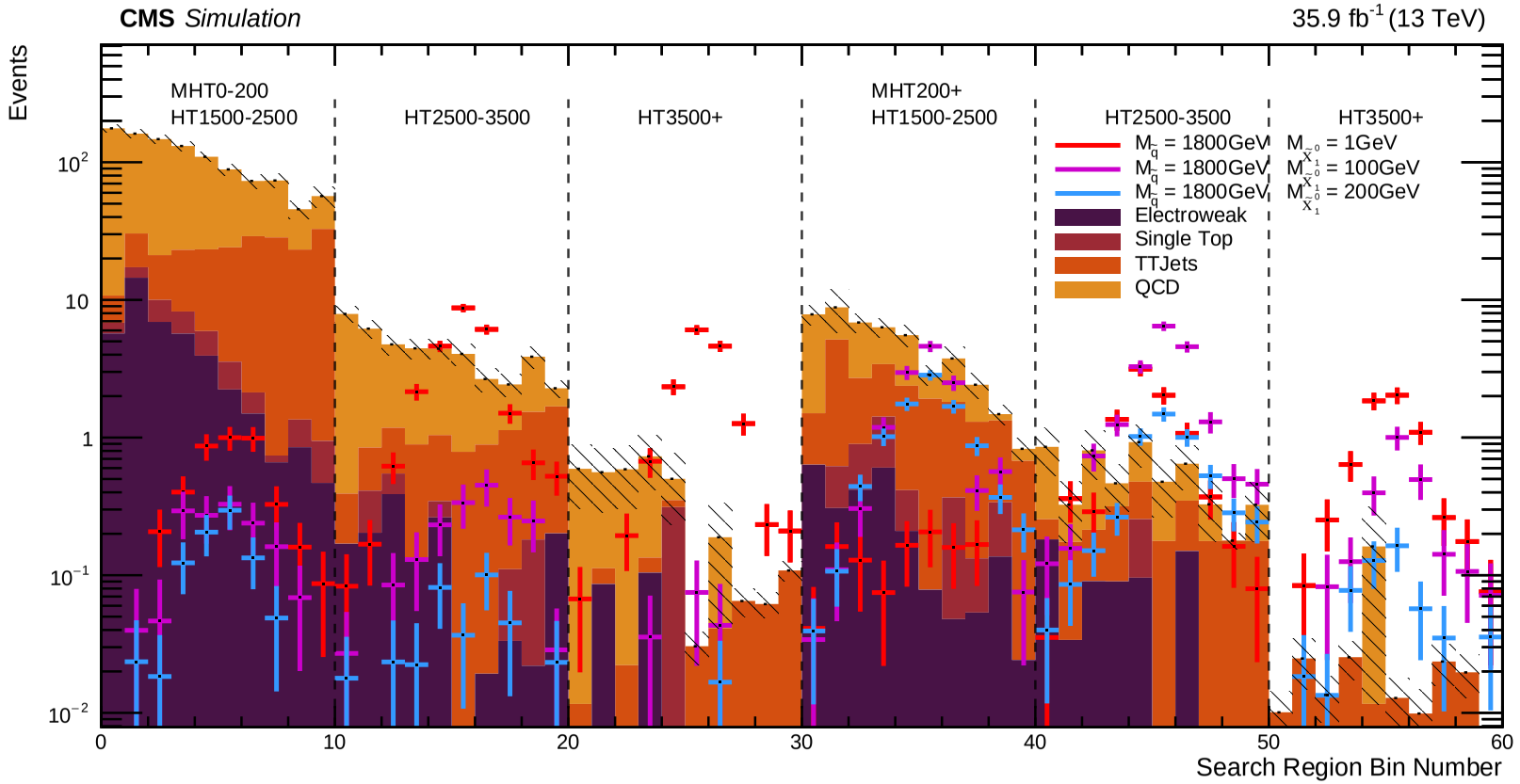


Figure D.2: Expected yields across the sixty measurement bins for three example signal mass points from the BM3 mass scan compared with SM MC background processes.

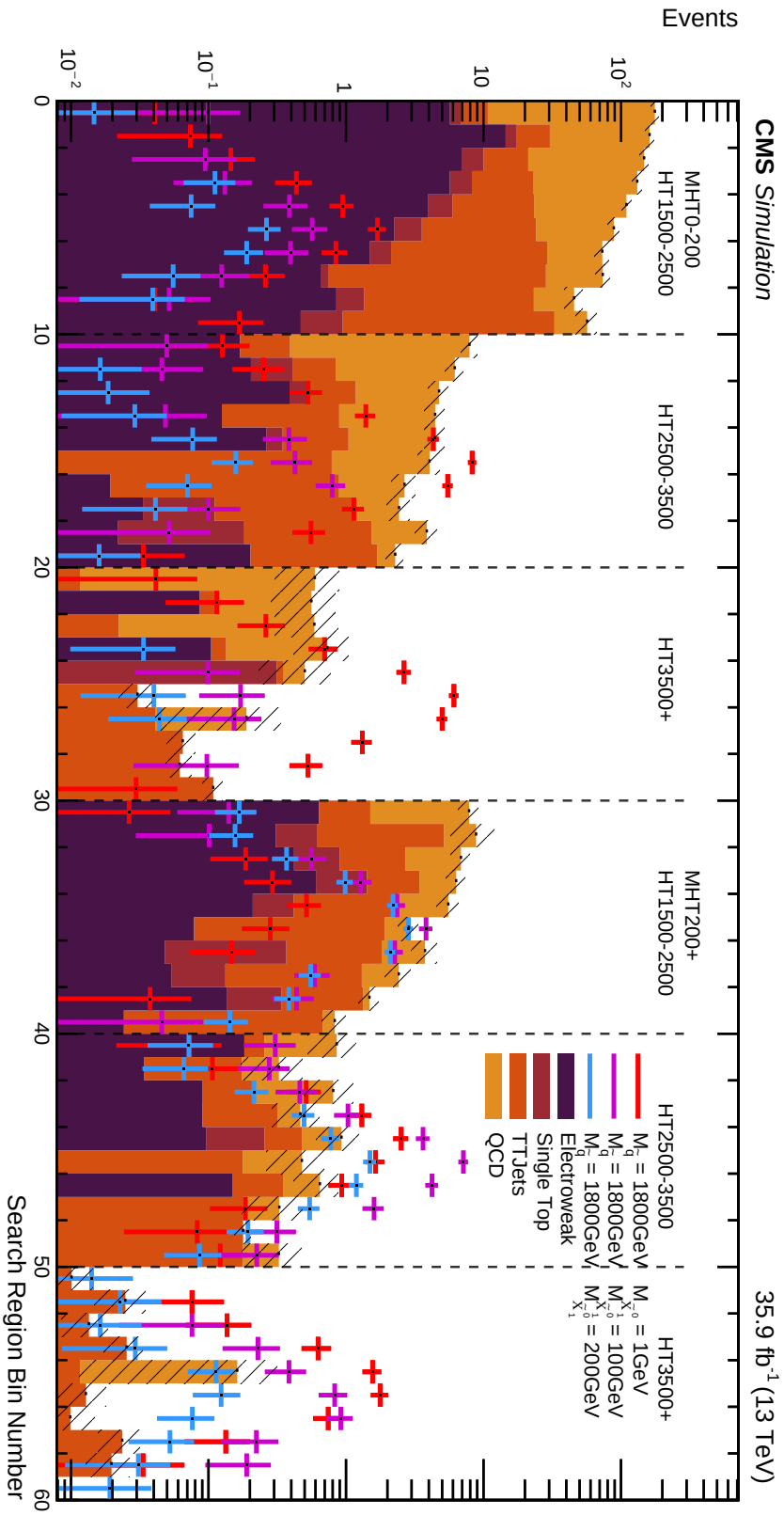


Figure D.3: Expected yields across the sixty measurement bins for three example signal mass points from the BM4 mass scan compared with SM MC background processes.

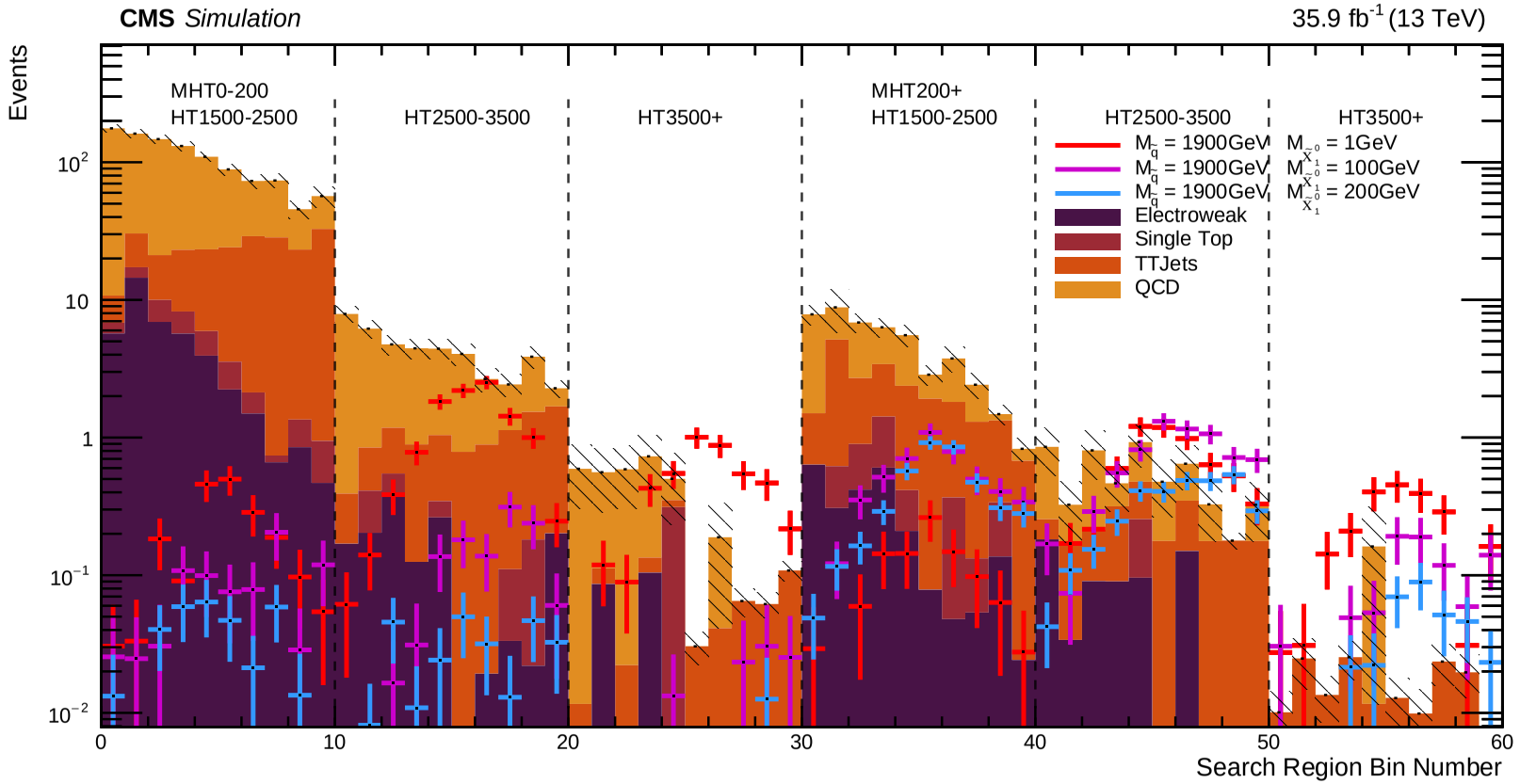


Figure D.4: Expected yields across the sixty measurement bins for three example signal mass points from the BM5 mass scan compared with SM MC background processes.

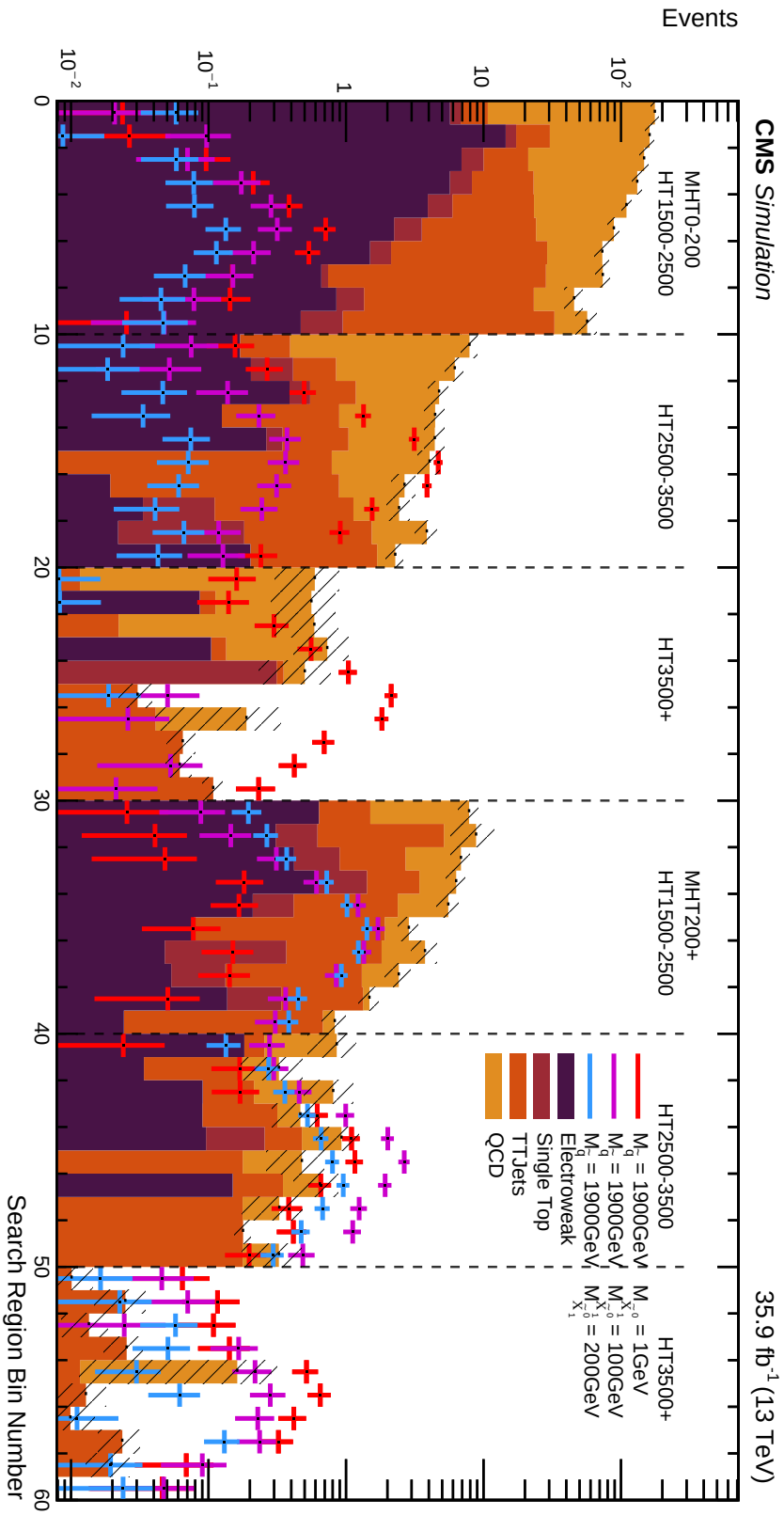


Figure D.5: Expected yields across the sixty measurement bins for three example signal mass points from the BM6 mass scan compared with SM MC background processes.

References

[1] A. Titterton et al., "Exploring Sensitivity to NMSSM Signatures with Low Missing Transverse Energy at the LHC", *JHEP* **10** (2018) 064, doi:10.1007/JHEP10(2018)064, arXiv:1807.10672.

[2] CMS Collaboration Collaboration, "A search for light Higgs bosons from supersymmetric decay cascades in proton-proton collisions at 13 TeV",.

[3] C.-N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance", *Phys. Rev.* **96** (1954) 191–195, doi:10.1103/PhysRev.96.191. [,150(1954)].

[4] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[5] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[6] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321. [,157(1964)].

[7] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508. [,160(1964)].

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585. [,162(1964)].

[9] K. Garrett and G. Duda, "Dark Matter: A Primer", *Adv. Astron.* **2011** (2011) 968283, doi:10.1155/2011/968283, arXiv:1006.2483.

[10] A. D. Sakharov, "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe", *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35, doi:10.1070/PU1991v034n05ABEH002497. [Usp. Fiz. Nauk161,no.5,61(1991)].

[11] M. Drewes, "The Phenomenology of Right Handed Neutrinos", *Int. J. Mod. Phys.* **E22** (2013) 1330019, doi:10.1142/S0218301313300191, arXiv:1303.6912.

[12] H. P. Nilles, "Supersymmetry, Supergravity and Particle Physics", *Phys. Rept.* **110** (1984) 1–162, doi:10.1016/0370-1573(84)90008-5.

[13] S. P. Martin, "A Supersymmetry primer", doi:10.1142/9789812839657_0001,10.1142/9789814307505_0001, arXiv:hep-ph/9709356. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[14] H. Georgi and S. L. Glashow, "Unity of All Elementary Particle Forces", *Phys. Rev. Lett.* **32** (1974) 438–441, doi:10.1103/PhysRevLett.32.438.

[15] CMS Collaboration, "Search for natural and split supersymmetry in proton-proton collisions at $\sqrt{s} = 13$ TeV in final states with jets and missing transverse momentum", *JHEP* **05** (2018) 025, doi:10.1007/JHEP05(2018)025, arXiv:1802.02110.

[16] R. B. Mann, "An introduction to particle physics and the standard model". CRC Press, Boca Raton, FL, 2010.

[17] J. Augustin et al., "Discovery of a narrow resonance in e+e- annihilation", *Physical Review Letters* **33** (1, 1974) 1406–1408, doi:10.1103/PhysRevLett.33.1406.

[18] M. L. Perl et al., "Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation", *Phys. Rev. Lett.* **35** (Dec, 1975) 1489–1492, doi:10.1103/PhysRevLett.35.1489.

[19] S. W. Herb et al., "Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions", *Phys. Rev. Lett.* **39** (1977) 252–255, doi:10.1103/PhysRevLett.39.252.

[20] UA1 Collaboration, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} =$ 540-GeV", *Phys. Lett.* **B122** (1983) 103–116, doi:10.1016/0370-2693(83)91177-2. [,611(1983)].

[21] UA1 Collaboration, "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c^2 at the CERN SPS Collider", *Phys. Lett.* **B126** (1983) 398–410, doi:10.1016/0370-2693(83)90188-0. [,7.55(1983)].

[22] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.

[23] D0 Collaboration, "Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. Lett.* **74** (1995) 2422–2426, doi:10.1103/PhysRevLett.74.2422, arXiv:hep-ex/9411001.

[24] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[25] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[26] I. J. R. Aitchison, "Supersymmetry in Particle Physics. An Elementary Introduction". Cambridge University Press, Cambridge, 2007.

[27] D. Bailin and A. Love, "Introduction to Gauge Field Theory Revised Edition". Graduate Student Series in Physics. Taylor & Francis, 1993.

[28] D. Griffiths, "Introduction to elementary particles". 2008.

[29] B. Bajc, J. Hisano, T. Kuwahara, and Y. Omura, "Threshold corrections to dimension-six proton decay operators in non-minimal SUSY SU (5) GUTs", *Nucl. Phys.* **B910** (2016) 1–22, doi:10.1016/j.nuclphysb.2016.06.017, arXiv:1603.03568.

[30] S. P. Martin, "Some simple criteria for gauged R-parity", *Phys. Rev.* **D46** (1992) R2769–R2772, doi:10.1103/PhysRevD.46.R2769, arXiv:hep-ph/9207218.

[31] S. Borgani, A. Masiero, and M. Yamaguchi, "Light gravitinos as mixed dark matter", *Phys. Lett.* **B386** (1996) 189–197, doi:10.1016/0370-2693(96)00956-2, arXiv:hep-ph/9605222.

[32] J. E. Kim and H. Nilles, "The Îij-problem and the strong CP-problem", *Physics Letters B* **138** (1984), no. 1, 150 – 154, doi:https://doi.org/10.1016/0370-2693(84)91890-2.

[33] OPAL Collaboration, "Search for chargino and neutralino production at $\sqrt{s}$ = 192-GeV to 209 GeV at LEP", *Eur. Phys. J.* **C35** (2004) 1–20, doi:10.1140/epjc/s2004-01758-8, arXiv:hep-ex/0401026.

[34] U. Ellwanger, C. Hugonie, and A. M. Teixeira, "The Next-to-Minimal Supersymmetric Standard Model", *Phys. Rept.* **496** (2010) 1–77, doi:10.1016/j.physrep.2010.07.001, arXiv:0910.1785.

[35] G. Giudice and A. Masiero, "A natural solution to the Îij-problem in supergravity theories", *Physics Letters B* **206** (1988), no. 3, 480 – 484, doi:https://doi.org/10.1016/0370-2693(88)91613-9.

[36] H.-S. Lee, K. T. Matchev, and T. T. Wang, "A U(1) -prime solution to the μ^- problem and the proton decay problem in supersymmetry without R-parity", *Phys. Rev.* **D77** (2008) 015016, doi:10.1103/PhysRevD.77.015016, arXiv:0709.0763.

[37] U. Ellwanger and A. M. Teixeira, "Excessive Higgs pair production with little MET from squarks and gluinos in the NMSSM", *JHEP* **04** (2015) 172, doi:10.1007/JHEP04(2015)172, arXiv:1412.6394.

[38] S. Moretti and S. Khalil, "Supersymmetry beyond minimality: from theory to experiment". CRC Press, 2017.

[39] I. Melzer-Pellmann and P. Pralavorio, "Lessons for SUSY from the LHC after the first run", *Eur. Phys. J. C* **74** (Apr, 2014) 2801. 31 p, doi:10.1140/epjc/s10052-014-2801-y.

[40] M. Maniatis, "The Next-to-Minimal Supersymmetric extension of the Standard Model reviewed", *Int. J. Mod. Phys.* **A25** (2010) 3505–3602, doi:10.1142/S0217751X10049827, arXiv:0906.0777.

[41] Yu. A. Golfand and E. P. Likhtman, "Extension of the Algebra of Poincare Group Generators and Violation of p Invariance", *JETP Lett.* **13** (1971) 323–326. [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)].

[42] J. Wess and B. Zumino, "Supergauge Transformations in Four-Dimensions", *Nucl. Phys.* **B70** (1974) 39–50, doi:10.1016/0550-3213(74)90355-1. [,24(1974)].

[43] R. Barbieri, S. Ferrara, and C. A. Savoy, "Gauge Models with Spontaneously Broken Local Supersymmetry", *Phys. Lett.* **119B** (1982) 343, doi:10.1016/0370-2693(82)90685-2.

[44] H. P. Nilles, "Supersymmetry, Supergravity and Particle Physics", *Phys. Rept.* **110** (1984) 1–162, doi:10.1016/0370-1573(84)90008-5.

[45] L. Randall and D. Tucker-Smith, "Dijet Searches for Supersymmetry at the LHC", *Phys. Rev. Lett.* **101** (2008) 221803, doi:10.1103/PhysRevLett.101.221803, arXiv:0806.1049.

[46] CMS Collaboration, "Search strategy for exclusive multi-jet events from supersymmetry at CMS",.

[47] ATLAS Collaboration, "Search for squarks and gluinos in final states with jets and missing transverse momentum using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector", *Phys. Rev.* **D97** (2018), no. 11, 112001, doi:10.1103/PhysRevD.97.112001, arXiv:1712.02332.

[48] ATLAS Collaboration, "Search for supersymmetry in final states with missing transverse momentum and multiple b-jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **06** (2018) 107, doi:10.1007/JHEP06(2018)107, arXiv:1711.01901.

[49] ATLAS Collaboration, "Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at $\sqrt{s}$=13 TeV with the ATLAS detector", *JHEP* **12** (2017) 085, doi:10.1007/JHEP12(2017)085, arXiv:1709.04183.

[50] CMS Collaboration, "Search for new phenomena with the M_{T2} variable in the all-hadronic final state produced in protonâĂŞproton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J.* **C77** (2017), no. 10, 710, doi:10.1140/epjc/s10052-017-5267-x, arXiv:1705.04650.

[51] CMS Collaboration, "Search for supersymmetry in multijet events with missing transverse momentum in proton-proton collisions at 13 TeV", *Phys. Rev.* **D96** (2017), no. 3, 032003, doi:10.1103/PhysRevD.96.032003, arXiv:1704.07781.

[52] CMS Collaboration, "Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton

collisions at $\sqrt{s} = 13$ TeV", *JHEP* **10** (2017) 005, doi:10.1007/JHEP10(2017)005, arXiv:1707.03316.

[53] M. Drees et al., "CheckMATE: Confronting your Favourite New Physics Model with LHC Data", *Comput. Phys. Commun.* **187** (2015) 227–265, doi:10.1016/j.cpc.2014.10.018, arXiv:1312.2591.

[54] D. Das, U. Ellwanger, and A. M. Teixeira, "NMSDECAY: A Fortran Code for Supersymmetric Particle Decays in the Next-to-Minimal Supersymmetric Standard Model", *Comput. Phys. Commun.* **183** (2012) 774–779, doi:10.1016/j.cpc.2011.11.021, arXiv:1106.5633.

[55] M. Muhlleitner, A. Djouadi, and Y. Mambrini, "SDECAY: A Fortran code for the decays of the supersymmetric particles in the MSSM", *Comput. Phys. Commun.* **168** (2005) 46–70, doi:10.1016/j.cpc.2005.01.012, arXiv:hep-ph/0311167.

[56] T. C. Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (aug, 2008) S08004–S08004, doi:10.1088/1748-0221/3/08/s08004.

[57] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[58] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[59] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[60] Linac4 Collaboration, A. Lombardi, "LINAC4 Commissioning", in *Proceedings, 29th International Linear Accelerator Conference (LINAC18): Beijing, China, September 16-21, 2018*, p. TH1P01. 2018. doi:10.18429/JACoW-LINAC2018-TH1P01.

[61] STFC, "CERN Accelerator Complex". http://www.stfc.ac.uk/research/particle-physics-and-particle-astrophysics/large-hadron-collider/cern-accelerator-complex/, 2006.

[62] CMS Collaboration, T. Sakuma, "3D SketchUp images of the CMS detector (120918)". http://cds.cern.ch/record/2628527, May, 2016.

[63] CMS Collaboration, D. Barney, "CMS Slice". https://cds.cern.ch/record/2628641, Feb, 2015.

[64] Q. Ingram, "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *Journal of Instrumentation* **2** (apr, 2007) P04004–P04004, doi:10.1088/1748-0221/2/04/p04004.

[65] CMS HCAL/ECAL Collaborations, "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c", *The European Physical Journal C* **60** (Apr, 2009) 359–373, doi:10.1140/epjc/s10052-009-0959-5.

[66] CMS Collaboration, "CMS Physics", *CERN-LHCC-2006-001, CMS-TDR-8-1* (2006).

[67] N. Bronson, T. Lento, and J. L. Wiener, "Open data challenges at Facebook", pp. 1516–1519. 04, 2015. doi:10.1109/ICDE.2015.7113415.

[68] V. Halyo et al., "GPU Enhancement of the Trigger to Extend Physics Reach at the LHC", *JINST* **8** (May, 2013) P10005. 13 p, doi:10.1088/1748-0221/8/10/P10005. Comments: 13 pages 8 figures.

[69] CMS Collaboration, "The CMS Level-1 electron and photon trigger: for Run II of LHC", *JINST* **12** (2017), no. 02, C02014, doi:10.1088/1748-0221/12/02/C02014.

[70] CMS collaboration Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", CERN-LHCC-2013-011. CMS-TDR-12, (Jun, 2013).

[71] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[72] A. Belyaev, N. D. Christensen, and A. Pukhov, "CalcHEP 3.4 for collider physics within and beyond the Standard Model", *Comput. Phys. Commun.* **184** (2013) 1729–1769, doi:10.1016/j.cpc.2013.01.014, arXiv:1207.6082.

[73] C. Oleari, "The POWHEG-BOX", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 36–41, doi:10.1016/j.nuclphysbps.2010.08.016, arXiv:1007.3893.

[74] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J.* **C58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[75] T. Gleisberg et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.

[76] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[77] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[78] B. Andersson, G. Gustafson, G. Ingelman, and T. SjÃűstrand, "Parton fragmentation and string dynamics", *Physics Reports* **97** (1983), no. 2, 31 – 145, doi:https://doi.org/10.1016/0370-1573(83)90080-7.

[79] A. Kupco, "Cluster hadronization in HERWIG 5.9", in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pp. 292–300. 1998. arXiv:hep-ph/9906412.

[80] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

[81] DELPHES 3 Collaboration, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", *JHEP* **02** (2014) 057, doi:10.1007/JHEP02(2014)057, arXiv:1307.6346.

[82] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[83] J. Allison et al., "Geant4 developments and applications", *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.

[84] J. Allison et al., "Recent developments in Geant4", *Nucl. Instrum. Meth.* **A835** (2016) 186–225, doi:10.1016/j.nima.2016.06.125.

[85] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[86] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[87] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup Per Particle Identification", *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[88] C. Weiser, "A Combined Secondary Vertex Based B-Tagging Algorithm in CMS", CMS-NOTE-2006-014, CERN, Geneva, (Jan, 2006).

[89] CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2",.

[90] CMS Collaboration, "Identification of double-b quark jets in boosted event topologies",.

[91] D. Krohn, M. D. Schwartz, M. Low, and L.-T. Wang, "Jet Cleansing: Pileup Removal at High Luminosity", *Phys. Rev.* **D90** (2014), no. 6, 065020, doi:10.1103/PhysRevD.90.065020, arXiv:1309.4777.

[92] CMS Collaboration, "Jet Identification for the 13 TeV data Run2016". https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID13TeVRun2016. Accessed 2019-05-08.

[93] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.

[94] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.

[95] CMS Collaboration, "Cut Based Electron ID for Run 2". https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2. Accessed 2019-06-12.

[96] CMS Collaboration, "Baseline muon selections for Run-II". https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2. Accessed 2019-06-12.

[97] W. Beenakker, R. Hopker, and M. Spira, "PROSPINO: A Program for the production of supersymmetric particles in next-to-leading order QCD", arXiv:hep-ph/9611232.

[98] W. Beenakker, R. Hopker, M. Spira, and P. M. Zerwas, "Squark and gluino production at hadron colliders", *Nucl. Phys.* **B492** (1997) 51–103, doi:10.1016/S0550-3213(97)80027-2, arXiv:hep-ph/9610490.

[99] C. Borschensky et al., "Squark and gluino production cross sections in pp collisions at $\sqrt{s}$ = 13, 14, 33 and 100 TeV", *Eur. Phys. J.* **C74** (2014), no. 12, 3174, doi:10.1140/epjc/s10052-014-3174-y, arXiv:1407.5066.

[100] CMS Collaboration, "Interpretation of Searches for Supersymmetry with Simplified Models", *Phys. Rev.* **D88** (2013), no. 5, 052017, doi:10.1103/PhysRevD.88.052017, arXiv:1301.2175.

[101] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].

[102] A. L. Read, "Presentation of search results: theCLstechnique", *Journal of Physics G: Nuclear and Particle Physics* **28** (sep, 2002) 2693–2704, doi:10.1088/0954-3899/28/10/313.

[103] CMS Collaboration, "Usage of b/c Tag Objects for 13 TeV Data in 2016 and 80X MC". https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco. Accessed 2019-05-07.

[104] CMS Collaboration, "Reweighting recipe to emulate Level 1 ECAL prefiring". https://twiki.cern.ch/twiki/bin/viewauth/CMS/\T1\L1ECALPrefiringWeightRecipe. Accessed 2019-04-11.

[105] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period",.

[106] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$ = 13 TeV",.

[107] CMS Collaboration, "Top PAG: Cross Sections subgroup". https://twiki.cern.ch/twiki/bin/view/CMS/TopCrossSections. Accessed 2019-03-21.

[108] CMS Collaboration, "Top PAG: Cross Sections subgroup". https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting. Accessed 2019-05-29.

[109] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011", CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, (Aug, 2011).

[110] G. Apollinari et al., "High-Luminosity Large Hadron Collider (HL-LHC)", *CERN Yellow Rep. Monogr.* **4** (2017) 1–516, doi:10.23731/CYRM-2017-004.

[111] T. Golling, "Future Circular Collider Study (FCC)", in *Proceedings, 3rd Large Hadron Collider Physics Conference (LHCP 2015): St. Petersburg, Russia, August 31-September 5, 2015*, pp. 559–564, Kurchatov Institute. Kurchatov Institute, Gatchina, 2016.